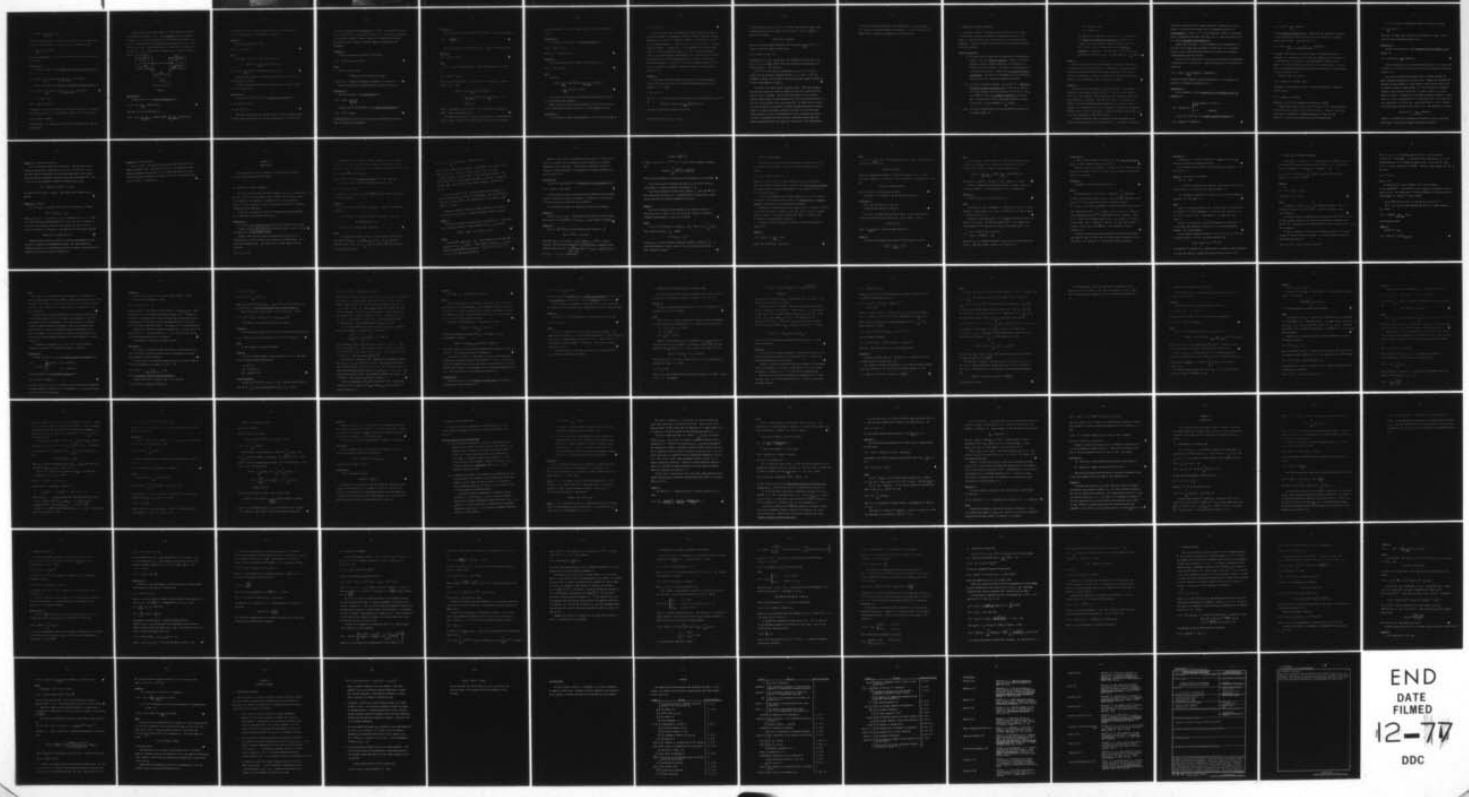
AD-A046 860 CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF COMPUTER --ETC F/G 9/2
GENERAL THEORY OF OPTIMAL ERROR ALGORITHMS AND ANALYTIC COMPLEX--ETC(U)
SEP 77 J F TRAUB, H WOZNIAKOWSKI N00014-76-C-0370

UNCLASSIFIED NL

1 OF
AD
A046860



END
DATE FILMED
12-77
DDC

AD A046860

(12) 2c

GENERAL THEORY OF OPTIMAL ERROR ALGORITHMS AND ANALYTIC COMPLEXITY

PART A. GENERAL INFORMATION MODEL

J. F. Traub
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

H. Woźniakowski
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213
(Visiting from the University of Warsaw)

DEPARTMENT
of
COMPUTER SCIENCE

DDC
RECEIVED
NOV 22 1977
B

AD No. ____
DDC FILE COPY

Carnegie-Mellon University

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

(6)

GENERAL THEORY OF OPTIMAL ERROR ALGORITHMS AND ANALYTIC COMPLEXITY.

PART A. GENERAL INFORMATION MODEL.

(9) Interim rept.,

(10) J. F. Traub
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

H. Woźniakowski
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213
(Visiting from the University of Warsaw)

(11) September 1977

(12) 95 p.

DDC
RECEIVED
NOV 22 1977
B

(15) This research was supported in part by the National Science Foundation under Grant MCS75-222-55 and the Office of Naval Research under Contract N00014-76-C-0370, NR 044-422.
NSF-MCS-75-22255

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

(17) AD A 043 081

LB

ABSTRACT

This is the first of a series of papers constructing an information based general theory of optimal errors and analytic computational complexity. Among the applications are such traditionally diverse areas as approximation, boundary-value problems, quadrature, and nonlinear equations in a finite or infinite dimensional space.

Traditionally algorithms are often derived by ad hoc criteria. The information based theory rationalizes the synthesis of algorithms by showing how to construct algorithms which minimize or nearly minimize the error. For certain classes of problems it shows how to construct algorithms (linear optimal error algorithms) which enjoy essentially optimal complexity with respect to all possible algorithms.

The existence of "strongly non-computable" problems is demonstrated. In contrast with the gap theorem of recursively computable functions we show that "every monotonic" real function is the complexity of some problem.

BY
DISTRIBUTION/AVAILABILITY
Dist.
A

TABLE OF CONTENTS

CHAPTER I

BASIC CONCEPTS

> Although this may seem a paradox, all exact science is dominated by the idea of approximation.
>
> B. Russell

1. INTRODUCTION

This is the first of a series of papers constructing an information based general theory of optimal error algorithms and analytic computational complexity. Among the applications are such traditionally diverse areas as approximation, boundary-value problems, quadrature, and nonlinear equations in a finite or infinite dimensional space.

In this paper we deal with "general information". Although general information can be used to solve nonlinear equations (see Section 15), "iterative information" is typically used for such problems. A theory of optimal error algorithms and optimal order for iterative information is developed in Traub and Woźniakowski [77c].

The general theory shows us how to construct algorithms which minimize or nearly minimize the error. For certain classes of problems we show how to construct algorithms (linear optimal error algorithms) which enjoy essentially optimal complexity with respect to all possible algorithms. The fact that optimal algorithms depend only on the "information" used permits a great simplification.

Traditionally, algorithms are derived by ad hoc criteria. The information based theory rationalizes the synthesis of algorithms. For illustration we consider a relatively simple problem-univariate numerical integration.

Gauss quadrature formulas are widely used. The formula is derived by the rather arbitrary criterion that it be exact for all polynomials of as

high a degree as possible. There is no reason to think this optimizes either error or complexity. The information based theory yields both optimal error algorithms and the fact that these algorithms are essentially optimal with respect to complexity among all possible algorithms. It turns out that the optimal error algorithms use equi-spaced abscissas and are therefore not Gaussian formulas. Furthermore their complexity varies inversely as the "regularity" of the class of integrands. It has been observed that the numerical integration problem cannot be solved to within pre-assigned error. We show (Section 12) that if we mildly restrict the class of integrands we can always solve the integration problem to within pre-assigned error.

The analysis needed to characterize and construct optimal error algorithms for a particular problem area can be difficult. Note, however, that this may be viewed as pre-conditioning since it need be done only once.

Among the major questions we pose and at least partially answer are:

1. What is a lower bound on the error of any algorithm for solving a problem using given information? See Section 2.

2. In general is there an algorithm which gets arbitrarily close to this lower bound? See Section 2.

3. When is the information strong enough to solve a problem arbitrarily closely? See Section 3.

4. What is the most relevant information for solving a problem? See Section 6.

5. If we are willing to pay enough can we always solve a problem approximately? See Sections 7 and 9.

6. For linear problems is there always a linear algorithm whose error is within a constant factor of having optimal error? See Section 7.

7. If we assume a Hilbert space setting, what characterizes those problems which can be solved arbitrarily closely? See Section 7.

8. For certain application areas how does optimal error and complexity depend on the "regularity" of the class of "problem elements"? See Sections 11, 12 and 13.

Analytic complexity is similar to abstract and concrete complexity in focusing on lower bounds of problem complexity. Analytic complexity obtains its characteristic flavor because:

1. Problems cannot be exactly solved with finite complexity.

2. In concrete complexity, problem elements are assumed given and information operators do not explicitly occur in the theory because they are identity operators. See Section 3 for examples. In analytic complexity only certain "information" about the problem element is obtainable.

We summarize some of our complexity results below.

1. Consideration of the pre-image set of an "information operator" gives us a general adversary principal for deriving lower bounds.

2. We demonstrate the existence of "strongly non-computable" problems. See Section 6.

3. We show the existence of arbitrarily hard problems. Furthermore "every monotonic" real function is the complexity of some problem. See Section 9. This may be contrasted with the gap theorem of recursively computable functions (Borodin [72]).

4. For certain classes of problems we can obtain lower bounds on the "inherent problem complexity" for all possible algorithms. See Sections 7 and 10.

5. For "linear optimal error algorithms" the difference between upper and lower bounds is very small. See, for example, Sections 9, 10, 11.

6. We perform worst case analysis over all problem elements in a class. In contrast with other recent complexity results, we believe that the worst case results hold for "almost all" problem elements.

Although we feel that progress has been made towards a general theory, much remains to be done. See Section 16 for a partial list of extensions. Application of the general theory to various problems of interest will require substantial work for each problem.

We summarize major concepts and results of this paper.

Section 2. We define $d(\mathfrak{N},S)$, the diameter of information $\mathfrak{N}$ for the problem S, and $r(\mathfrak{N},S)$, the radius of information $\mathfrak{N}$ for the problem S and show (Theorem 2.1) that $r(\mathfrak{N},S)$ provides a best possible lower bound on the error of any algorithm using the information $\mathfrak{N}$. We define interpolatory algorithm and show (Theorems 2.2, 2.3) that any interpolatory algorithm has an error which differs by at most a factor of two from a lower bound on the error.

We also show that there is always an algorithm which is arbitrarily close to the lower bound. We observe (Corollary 2.2) that a problem S using information $\mathfrak{N}$ can be solved to within an error whose norm is at most ϵ iff $r(\mathfrak{N},S) < \epsilon$.

Section 3. We introduce our model of computation and the major complexity concepts. In particular, we define primitive operations, permissible information operators and algorithms. We define the ϵ-complexity of a problem in a class of permissible information operators Υ as the complexity of solving problem S to within ϵ if the best algorithm and the best information from Υ are used.

Section 4. In Sections 4-9 we consider linear problems and linear information. The cardinality (card($\mathfrak{N}$)) of a linear information operator is defined and we show (Lemma 4.2) that information operators with finite cardinality equal to n can be represented by n linearly independent linear functionals.

Section 5. We consider problems specified by a linear solution operator S and a linear restriction operator T. We show that the dependence of $d(\mathfrak{N},S,T)$ on $\mathfrak{N}$ is only through the kernel of $\mathfrak{N}$. We define index(S,T) and show (Theorem 5.2) that if $\text{card}(\mathfrak{N}) < \text{index}(S,T)$ the solution cannot be approximated to within ϵ even for arbitrarily large ϵ. In particular, if index $(S,T) = \infty$, the problem cannot be solved by any information operator with finite cardinality.

Section 6. For fixed cardinality what is the most relevant information for solving a problem? Formally, the n-th minimal diameter of information, $d(n,S,T)$, is the diameter if the best information of cardinality at most n is used. Theorem 6.1 shows $d(n,S,T)$ is completely determined by the

operator ST^{-1}. A problem is ε-non-computable if $d(S,T) = \lim_{n\to\infty} d(n,S,T) \geq 2\varepsilon$ and is convergent if $d(S,T) = 0$. We show that $d(S,T)$ can be any number.

Section 7. If the image of the restriction operator is a Hilbert space then the problem (S,T) is convergent iff ST^{-1} is a compact operator. This implies (Corollary 7.1) the existence of linear problems which cannot be solved to within arbitrary ε with any finite number of linear functionals. In a Hilbert space the problem of most relevant information of cardinality n is completely solved (Theorem 7.2).

Section 8. To minimize combinatorial complexity it is desirable to use linear algorithms. We construct (Theorem 8.1) linear interpolatory algorithms whose error is within a numerical factor of optimality. If in particular the image of the restriction operator is a Hilbert space then we construct (Corollary 8.1) a linear interpolatory optimal error algorithm.

Section 9. We specify our model of computation for the linear case. We show (Theorem 9.2) there exist linear problems with essentially arbitrary complexity. This implies (Corollary 9.1) there exist arbitrarily hard linear problems and that there are no "gaps" in the complexity function.

Sections 10-15. We apply the general theory to a variety of problems. See the Table of Contents for a list. We confine ourselves here to relatively simple problems in order to concentrate on ideas and avoid overwhelming the reader with technical details.

Section 16. We list some extensions to the theory which will be reported elsewhere.

Glossary. For the reader's convenience a glossary of important definitions and symbols is provided.

2. DIAMETER AND RADIUS OF GENERAL INFORMATION

Let $\mathfrak{I}_0$ be a subset of a linear space $\mathfrak{I}_1$ over the real or complex field. Consider a linear or nonlinear operator S such that

(2.1) $S: \mathfrak{I}_0 \to \mathfrak{I}_2$

where $\mathfrak{I}_2$ is a linear normed space over the real or complex field. Let $\varepsilon > 0$ be a given number. Our problem is to find an <u>ε-approximation</u> $x = x(f)$, $x \in \mathfrak{I}_2$, to $\alpha = S(f)$, i.e.

(2.2) $\|x-\alpha\| < \varepsilon$

for all $f \in \mathfrak{I}_0$. We shall call S the <u>solution operator</u>, f a <u>problem element</u> and α a <u>solution element</u>. We shall often refer to S and its domain $\mathfrak{I}_0$ as the <u>problem S</u>.

To find an ε-approximation we must know something about the operator S. Let

(2.3) $\mathfrak{N}: D_{\mathfrak{N}} \to \mathfrak{I}_3$

be an <u>information operator</u> (not necessarily linear) where $\mathfrak{I}_0 \subset D_{\mathfrak{N}} \subset \mathfrak{I}_1$ and $\mathfrak{I}_3$ is a given space. $\mathfrak{N}(f)$ is called the <u>information of f</u>. For most problems the information operator $\mathfrak{N}$ is not one-to-one and $\mathfrak{N}(f)$ does not uniquely define the solution element $\alpha = S(f)$. Thus there may exist many different problem elements $f \in \mathfrak{I}_0$ with the same information.

Let $f \in \mathfrak{I}_0$. Let

(2.4) $V(f) = \{\tilde{f}: \mathfrak{N}(\tilde{f}) = \mathfrak{N}(f) \text{ and } \tilde{f} \in \mathfrak{I}_0\}$

be the pre-image set of $y = \mathfrak{N}(f)$, $V(f) = \mathfrak{N}^{-1}(y)$. Note that $V(f)$ is not empty since $f \in V(f)$ for every $f \in \mathfrak{I}_0$. Furthermore let

$$(2.5) \quad U(f) = \{S(\tilde{f}): \ \tilde{f} \in V(f)\}$$

be the set of all solutions $S(\tilde{f})$ of problem elements $\tilde{f}$ which share the same information as f, $U(f) = S(\mathfrak{N}^{-1}(y))$. Then knowing only $\mathfrak{N}(f)$ it is impossible to recognize which solution element $\alpha = S(f)$ or $\tilde{\alpha} = S(\tilde{f})$ is being actually approximated for all $\tilde{f} \in V(f)$. This adversary principal can be schematized as follows:

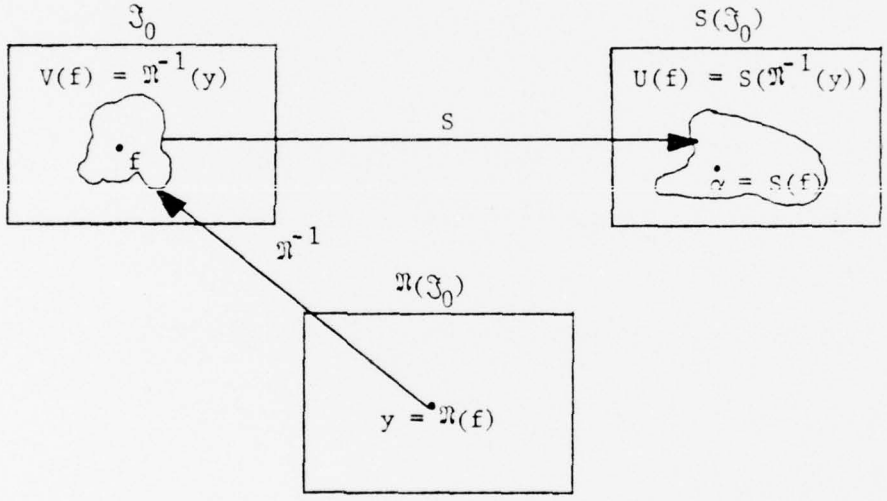


Figure 1

As we shall see below, the diameter $d(\mathfrak{N},S)$ and the radius $r(\mathfrak{N},S)$ of the set $U(f)$ play essential roles. These concepts are defined as follows.

Recall that for a set A, $A \subset \mathfrak{I}_2$,

$$(2.6) \quad \operatorname{diam}(A) = \sup_{a_1,a_2 \in A} \| a_1 - a_2 \|$$

is called the <u>diameter of A</u> and

$$(2.7)\quad \mathrm{rad}(A) = \inf_{a\in\mathfrak{Z}_2} \sup_{a_1\in A} \| a-a_1 \|$$

is called the radius of A. Roughly speaking, rad(A) is the minimal radius of a "ball" which contains A. If there exists c, $c \in \mathfrak{Z}_2$, such that

$$(2.8)\quad \sup_{a_1\in A} \| c-a_1 \| = \mathrm{rad}(A)$$

then c is a center of A. Note that c can be an element outside A and need not be unique.

Definition 2.1

We shall say $d(\mathfrak{N},S)$ is the diameter of information $\mathfrak{N}$ for the problem S iff

$$(2.9)\quad d(\mathfrak{N},S) = \sup_{f\in\mathfrak{Z}_0} \mathrm{diam}(U(f)) \left(= \sup_{f\in\mathfrak{Z}_0} \sup_{\tilde{f}\in V(f)} \| S(\tilde{f})-S(f) \| \right).$$

We shall say $r(\mathfrak{N},S)$ is the radius of information $\mathfrak{N}$ for the problem S iff

$$(2.10)\quad r(\mathfrak{N},S) = \sup_{f\in\mathfrak{Z}_0} \mathrm{rad}(U(f)) \left(= \sup_{f\in\mathfrak{Z}_0} \inf_{a\in\mathfrak{Z}_2} \sup_{\tilde{f}\in V(f)} \| a-S(\tilde{f}) \| \right). \quad \blacksquare$$

It is obvious that

$$(2.11)\quad d(\mathfrak{N},S) \le 2r(\mathfrak{N},S).$$

Furthermore if U(f) has a center c(f) for every $f \in \mathfrak{Z}_0$ and U(f) is symmetric with respect to c(f), i.e. $u + c(f) \in U(f)$ implies $-u + c(f) \in U(f)$, then

$$(2.12)\quad d(\mathfrak{N},S) = 2r(\mathfrak{N},S).$$

For many $\mathfrak{N}$ and S it is much easier to compute the diameter $d(\mathfrak{N},S)$ than the radius $r(\mathfrak{N},S)$.

We shall show that the radius $r(\mathfrak{N},S)$ is a lower bound on the error of any algorithm for solving $\alpha = S(f)$. By an algorithm we mean an operator φ: $\mathfrak{N}(\mathfrak{I}_0) \to \mathfrak{I}_2$. (See also the definition of "permissible algorithm" in Section 3.) We are interested in algorithms which approximate $\alpha = S(f)$. Let $\Phi(\mathfrak{N},S)$ be the class of all such algorithms. Since $\varphi(\mathfrak{N}(\tilde{f})) = \varphi(\mathfrak{N}(f))$ for all $\tilde{f} \in V(f)$, φ has to approximate any element of the set $U(f) = S(\mathfrak{N}^{-1}(y))$. This is shown in Figure 2.

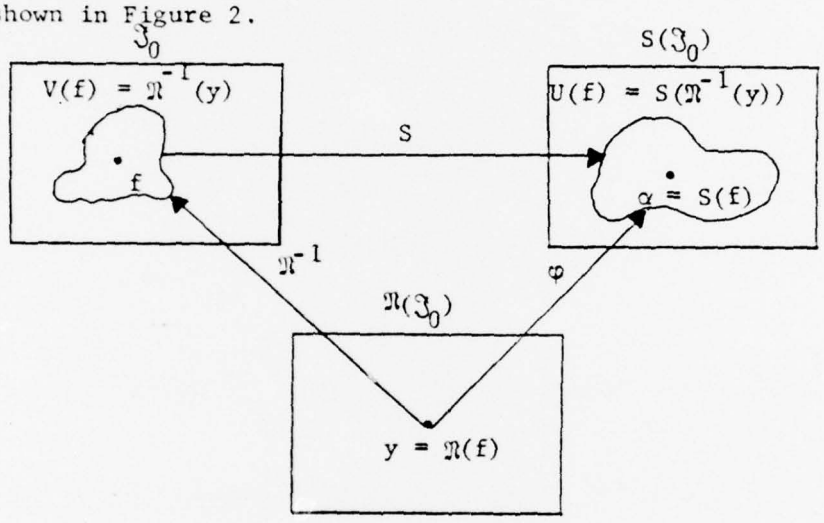


Figure 2

Definition 2.2

We shall say $e(\varphi)$ is the error of algorithm φ iff

$$(2.13)\quad e(\varphi) = \sup_{f \in \mathfrak{I}_0} \| \varphi(\mathfrak{N}(f)) - S(f) \| . \qquad \blacksquare$$

Note that (2.13) can be rewritten as

$$(2.14)\quad e(\varphi) = \sup_{f \in \mathfrak{I}_0} \sup_{\tilde{f} \in V(f)} \| \varphi(\mathfrak{N}(f)) - S(\tilde{f}) \| = \sup_{f \in \mathfrak{I}_0} \sup_{\tilde{\alpha} \in U(f)} \| \varphi(\mathfrak{N}(f)) - \tilde{\alpha} \| .$$

It is intuitively obvious that the radius $r(\mathfrak{N},S)$ is a lower bound on the error of any algorithm. A formal proof is provided by

Theorem 2.1

For any algorithm φ, $\varphi \in \Phi(\mathfrak{N},S)$,

$$e(\varphi) \geq r(\mathfrak{N},S). \tag{2.15}$$ ■

Proof

Let $f \in \mathfrak{F}_0$. Then due to (2.7) and (2.14) we get

$$\operatorname{rad}(U(f)) \leq \sup_{\tilde{\alpha} \in U(f)} \| \varphi(\mathfrak{N}(f)) - \tilde{\alpha} \| \leq e(\varphi).$$

Thus $r(\mathfrak{N},S) = \sup_{f \in \mathfrak{F}_0} \operatorname{rad}(U(f)) \leq e(\varphi)$ which proves (2.15). ■

This generalizes Theorem 4 in Micchelli and Rivlin [77] where S and $\mathfrak{N}$ are assumed to be linear.

We define "interpolatory algorithms" and show they are within a factor of 2 of the radius $r(\mathfrak{N},S)$.

Definition 2.3

An algorithm φ^I, $\varphi^I \in \Phi(\mathfrak{N},S)$, is an <u>interpolatory algorithm</u> iff

$$\varphi^I(\mathfrak{N}(f)) = S(\tilde{f}) \tag{2.16}$$

for some $\tilde{f} \in V(f)$. ■

This means that knowing the information $\mathfrak{N}(f)$ one finds a problem element $\tilde{f}$ (which always exists) which has the same information as f, $\tilde{f} \in V(f)$, and

$\tilde{\alpha} = S(\tilde{f})$ is proposed as an approximation to $\alpha = S(f)$. In practice $\tilde{f}$ is chosen to be "simpler" than f. Note that $\varphi^I(\mathfrak{N}(f)) \in U(f)$. In some cases, an assumption how to choose a unique $\tilde{f}$ is added. Examples of interpolatory algorithms are known for such problems as nonlinear equations, approximation and quadrature.

Theorem 2.2

For any interpolatory algorithm φ^I, $\varphi^I \in \Phi(\mathfrak{N},S)$,

$$(2.17) \quad e(\varphi^I) \leq d(\mathfrak{N},S) \leq 2\, r(\mathfrak{N},S). \qquad \blacksquare$$

Proof

Take any $f \in \mathfrak{I}_0$. Then

$$\| \varphi^I(\mathfrak{N}(f)) - S(f) \| \leq \| S(\tilde{f}) - S(f) \| \leq d(\mathfrak{N},S)$$

since $\tilde{f} \in V(f)$. Taking the supremum with respect to f we get (2.17). $\blacksquare$

We seek "optimal error algorithms" which are defined as follows.

Definition 2.4

We shall say $e(\mathfrak{N},S)$ is the optimal error iff

$$(2.18) \quad e(\mathfrak{N},S) = \inf_{\varphi \in \Phi(\mathfrak{N},S)} e(\varphi).$$

We shall say φ^{oe}, $\varphi^{oe} \in \Phi(\mathfrak{N},S)$, is an optimal error algorithm iff

$$(2.19) \quad e(\varphi^{oe}) = e(\mathfrak{N},S). \qquad \blacksquare$$

Combining Theorems 2.1 and 2.2 we see that any interpolatory algorithm is nearly an optimal error algorithm.

Corollary 2.1

For any interpolatory algorithm φ^I, $\varphi^I \in \mathfrak{I}(\mathfrak{N},S)$, with the convention $\frac{0}{0} = 1$,

$$(2.20)\quad 1 \le \frac{e(\varphi^I)}{e(\mathfrak{N},S)} \le 2. \qquad \blacksquare$$

We now prove that the optimal error $e(\mathfrak{N},S)$ is equal to the radius $r(\mathfrak{N},S)$.

Theorem 2.3

$$(2.21)\quad e(\mathfrak{N},S) = r(\mathfrak{N},S). \qquad \blacksquare$$

Proof

Let $\delta > 0$ be an arbitrary number. Define an algorithm φ_δ as follows. Let

$$(2.22)\quad \varphi_\delta(\mathfrak{N}(f)) = c_\delta(f)$$

where $\| c_\delta(f) - \tilde{\alpha} \| \le \mathrm{rad}(U(f)) + \delta$ for all $\tilde{\alpha} \in U(f)$. Thus $c_\delta(f)$ is almost a center of $U(f)$. Then

$$e(\mathfrak{N},S) \le e(\varphi_\delta) = \sup_{f \in \mathfrak{F}_0} \| S(f) - \varphi_\delta(\mathfrak{N}(f)) \| =$$

$$= \sup_{f \in \mathfrak{F}_0} \sup_{\tilde{\alpha} \in U(f)} \| \tilde{\alpha} - c_\delta(f) \| \le \sup_{f \in \mathfrak{F}_0} \mathrm{rad}(U(f)) + \delta =$$

$$= r(\mathfrak{N},S) + \delta.$$

Since δ is arbitrary, $e(\mathfrak{N},S) \le r(\mathfrak{N},S)$. Due to Theorem 2.1 we know that $e(\mathfrak{N},S) \ge r(\mathfrak{N},S)$ which proves (2.22). $\blacksquare$

See Micchelli and Rivlin [77] where a similar result is established for $\mathfrak{N}$ and S linear. Theorem 2.3 motivates using a center c(f), if it exists,

as an approximation to $\alpha = S(f)$. Suppose that $U(f)$ has a center for any $f \in \mathfrak{F}_0$.

Definition 2.5

An algorithm φ^c, $\varphi^c \in \Phi(\mathfrak{N},S)$ is a central algorithm iff

$$\varphi^c(\mathfrak{N}(f)) = c(f) \tag{2.23}$$

where $c(f)$ is a center of $U(f)$. ■

Theorem 2.4

Any central algorithm is an optimal error algorithm, i.e.,

$$e(\varphi^c) = r(\mathfrak{N},S). \tag{2.24}$$ ■

Proof

Note that

$$e(\varphi^c) = \sup_{f\in\mathfrak{F}_0} \| S(f)-\varphi^c(\mathfrak{N}(f)) \| = \sup_{f\in\mathfrak{F}_0} \sup_{\tilde{\alpha}\in U(f)} \| \tilde{\alpha}-c(f) \|$$

$$= \sup_{f\in\mathfrak{F}_0} \operatorname{rad}(U(f)) = r(\mathfrak{N},S).$$ ■

As we shall see in Section 8 an interpolatory algorithm may turn out to be an optimal error algorithm.

Recall we wish to find an ε-approximation to $\alpha = S(f)$ for all $f \in \mathfrak{F}_0$, i.e., to find $x(f)$ such that $\| x(f)-\alpha \| < \varepsilon$. Due to Theorem 2.3 we get

Corollary 2.2

It is possible to find an ε-approximation to $\alpha = S(f)$ for all $f \in \mathfrak{F}_0$ iff

(2.25) $r(\mathfrak{N},S) < \varepsilon$. ■

We wish to stress that an information operator $\mathfrak{N}$ has to be defined in such a way that $\mathfrak{N}(f)$ is "computable" for every $f \in \mathfrak{I}_0$. This rules out many operators as "permissible" information operators. For instance, let $\mathfrak{N}(f) = f$ be the identity operator I. Then, since I is one to one, $r(I,S) = 0$ for any solution operator S. However $\mathfrak{N}(f) = f$ is "computable" iff f can be represented by a finite dimensional vector, i.e., $\mathfrak{I}_1$ is a finite dimensional space. As a second example, consider $\mathfrak{N}(f) = S(f)$. Then $r(S,S) = 0$ but for most problems S(f) is not "computable" and $\mathfrak{N} = S$ is not a "permissible" information operator. See Sections 3 and 9 for a precise definition of our model of computation. Examples of computable operators will be found in Sections 10-15.

Example 2.1

To illustrate the above concepts we consider the following problem. Let $\mathfrak{I}_1 = C^n[0,1]$ be the class of n times differentiable functions of one variable, $n \geq 1$. Define

(2.25) $S(f) = f$

that is, S = I. Note that (2.25) is a formulation of the approximation problem. Let

$$f \in \mathfrak{I}_0 = \{f: f \in \mathfrak{I}_1 \text{ and } \max_{0 \leq t \leq 1} |\frac{1}{n!} f^{(n)}(x)| \leq 1\}.$$

Consider the information operator $\mathfrak{N}$ given by

(2.26) $\mathfrak{N}(f) = [f(t_1), f(t_2), \ldots, f(t_n)]$

for some distinct points $t_i \in [0,1]$. This means that we want to approximate f from $\mathfrak{I}_0$ knowing only the values of f at n points. Let $\omega(t) = \prod_{i=1}^{n} (t-t_i)$. Then $\tilde{f} \in V(f)$ implies

$$\tilde{f}(t) = f(t) + g(t)\omega(t)$$

where g is the nth divided difference of $\tilde{f}$-f and $\|g\| = \max_{0\le t\le 1} |g(t)| \le 2$. It is easy to show that $d(\mathfrak{N},S) = 2r(\mathfrak{N},S)$ and

(2.27) $\quad r(\mathfrak{N},S) = \|\omega\| \ge 2/4^n$.

Furthermore (2.27) holds with equality for information $\mathfrak{N}$ of the form (2.26) with $t_i = t_i^* = 2\cos(\frac{\pi}{2n} + \frac{i-1}{n}\pi)$ which are the zeros of the Chebyshev polynomial $T_n(\frac{1}{2}(t+1))$. See Section 11.

If $\varepsilon \ge 2/4^n$ we can find an ε-approximation to $\alpha = S(f) = f$ for all $f \in \mathfrak{I}_0$ using the information operator $\mathfrak{N}$ with $t_i = t_i^*$. For $\varepsilon < 2/4^n$, the information operator $\mathfrak{N}$ of the form (2.26) does not supply enough information to find ε-approximations for <u>any</u> t_i. ■

We conclude this section with a historical note. The ideas presented here have been implicitly used by a number of people for a particular problem or a class of problems. The key point was always to find a problem element $\tilde{f}$ which shares the same information as f and the distance between $S(f)$ and $S(\tilde{f})$ was an inherent error of any algorithm. See among others Winograd [76] who introduced a very general "fooling" technique and showed its importance for a number of problems, Micchelli and Rivlin [77] who considered linear operators S and $\mathfrak{N}$, and Woźniakowski [75] who introduced the concept of order of information for the solution of nonlinear operator equations. Brent, Winograd and Wolfe [74], Kung [75], Kacewicz [75, 76a, 76b], Meersman

[76a,76b], Traub and Woźniakowski [76a], Woźniakowski [76] all considered the solution of nonlinear equations and Werschulz [77a,77b] dealt with the maximal order of numerical integration and differentiation.

3. COMPLEXITY OF GENERAL INFORMATION

We present our model of computation which consists of a set of primitive operations, permissible information operators and permissible algorithms. In what follows we shall use the words cost and complexity interchangeably. Context will make it clear whether we mean algorithm complexity or problem complexity.

Model of Computation

(i) We assume that the computations are performed on a random access machine. Let p be a primitive operation. Examples of primitive operations include arithmetic operations, the evaluation of a square root or of an integral. Let comp(p) be the complexity of p; comp(p) must be finite. Suppose that P is a given collection of primitives. The choice of P and comp(p), $p \in P$, are arbitrary and can depend on the particular problem being solved.

(ii) Let $\mathfrak{N}$ be an information operator. We say that $\mathfrak{N}$ is a permissible information operator with respect to P if $\mathfrak{N}(f)$ can be computed by a finite number of primitive operations from P for all $f \in \mathfrak{F}_0$. Let $\mathrm{comp}(\mathfrak{N}(f))$ denote the information complexity of computing $\mathfrak{N}(f)$. We assume that if $\mathfrak{N}(f)$ requires the evaluation of primitives $p_1, p_2, \ldots, p_k$ then $\mathrm{comp}(\mathfrak{N}(f)) = \sum_{i=1}^{k} \mathrm{comp}(p_i)$.

(iii) Let φ be an algorithm which uses the permissible information $\mathfrak{N}$. To evaluate $\varphi(\mathfrak{N}(f))$ we:

(a) compute $y = \mathfrak{N}(f)$,

(b) compute $\varphi(y)$.

The complexity of computing y is given by (ii). We say that φ is a permissible algorithm with respect to P if $\varphi(y)$ can be computed by a finite number of primitive operations from P for all $y = \mathfrak{N}(f)$, $f \in \mathfrak{I}_0$. Let $\text{comp}(\varphi(y))$ be the combinatory complexity of computing y. We assume that if $\varphi(y)$ requires the evaluation of primitives $q_1, q_2, \ldots, q_j$ then $\text{comp}(\varphi(y)) = \sum_{i=1}^{j} \text{comp}(q_i)$. ■

Remark 3.1

Let $\mathfrak{N}$ be a permissible information operator. This means that $\mathfrak{N}(f)$ can be computed from the set of primitives P. Often there exist many different algorithms for computing $\mathfrak{N}(f)$ and the optimal computation of $\mathfrak{N}(f)$ can be treated as a subproblem. However, we assume that an algorithm (possibly not optimal) for the computation of $\mathfrak{N}(f)$ is defined by a "user". ■

Example 3.1

Suppose we wish to approximate $S(f) = \int_0^1 f(t)dt$ where f' is a piecewise continuous scalar function and $\int_0^1 (f'(t))^2 dt \leq 1$. Define two sets of primitives, P_1 = {the evaluation of an integral} and P_2 = {arithmetic operations, the evaluation of a function}. Note that $\mathfrak{N}(f) = S(f)$ is permissible with respect to P_1 and not permissible with respect to P_2. Of course $r(S,S) = 0$. However, S is a primitive only in P_1. An example of a permissible information operator for P_2 is $\mathfrak{N}(f) = [f(t_1), f(t_2), \ldots, f(t_n)]$ for equally spaced $t_i \in [0,1]$. It is shown in Section 12 that $r(\mathfrak{N},S) = O(1/n)$. ■

We showed in Section 2 that a necessary and sufficient condition for finding an ε-approximation to $\alpha = S(f)$ is $r(\mathfrak{N},S) < \varepsilon$. If $r(\mathfrak{N},S) \geq \varepsilon$ then the

information operator $\mathfrak{N}$ does not supply sufficient information to solve the problem. We say that the problem S with an information operator $\mathfrak{N}$ is ε-non-computable if $r(\mathfrak{N},S) \geq \varepsilon$. If $\mathfrak{N}$ is permissible, $r(\mathfrak{N},S) < \varepsilon$, and there exists a permissible algorithm φ such that $e(\varphi) < \varepsilon$, then the problem S with $\mathfrak{N}$ is called ε-computable with respect to P.

Suppose then that $r(\mathfrak{N},S) < \varepsilon$ for a permissible $\mathfrak{N}$ and assume that $\Phi(\varepsilon)$, the class of all permissible algorithms for which $e(\varphi) < \varepsilon$, is non-empty. We want to derive lower and upper bounds on the complexity of finding an ε-approximation using any $\varphi \in \Phi(\varepsilon)$.

Since the set of primitives P is fixed, we do not mention the dependence of complexity on P. Let $\varphi \in \Phi(\varepsilon)$. Then the complexity of an algorithm φ is defined by

$$(3.1) \quad \mathrm{comp}(\varphi) = \sup_{f \in \mathfrak{I}_0} (\mathrm{comp}(\mathfrak{N}(f)) + \mathrm{comp}(\varphi(\mathfrak{N}(f)))).$$

We define optimal complexity algorithm (Definition 3.1) and inherent complexity of a problem (see Definition 3.2).

Definition 3.1

We say $\mathrm{comp}(\mathfrak{N},S,\varepsilon)$ is the ε-complexity of the information $\mathfrak{N}$ for the problem S iff

$$(3.2) \quad \mathrm{comp}(\mathfrak{N},S,\varepsilon) = \begin{cases} \inf_{\varphi \in \Phi(\varepsilon)} \mathrm{comp}(\varphi), & \text{if } r(\mathfrak{N},S) < \varepsilon \\ +\infty, & \text{otherwise} \end{cases}$$

We say φ^{oc}, $\varphi^{oc} \in \Phi(\varepsilon)$ is an optimal complexity algorithm iff

$$(3.3) \quad \mathrm{comp}(\varphi^{oc}) = \mathrm{comp}(\mathfrak{N},S,\varepsilon). \qquad \blacksquare$$

Let

$$(3.4)\quad \mathrm{comp}(\mathfrak{N}) = \sup_{f \in \mathfrak{F}_0} \mathrm{comp}(\mathfrak{N}(f))$$

be the information complexity of $\mathfrak{N}$. Suppose that the combinatory complexity of every algorithm φ, $\varphi \in \Phi(\varepsilon)$, for $\mathfrak{N}(f)$ such that $\mathrm{comp}(\mathfrak{N}(f)) = \mathrm{comp}(\mathfrak{N})$ is bounded below by $m(\mathfrak{N})$. More precisely, let

$$(3.5)\quad m(\mathfrak{N}) = \inf_{\varphi \in \Phi(\varepsilon)} \sup_{f:\ \mathrm{comp}(\mathfrak{N}(f)) = \mathrm{comp}(\mathfrak{N})} \mathrm{comp}(\varphi(\mathfrak{N}(f))).$$

In general, $m(\mathfrak{N})$ depends on the total number of "independent pieces" of information $\mathfrak{N}$. See Section 4 and 9 where the "cardinality" of information $\mathfrak{N}$ is introduced and its influence on the combinatorial complexity of φ is shown. For linear problems, as we shall see in Sections 8 and 9, it is possible to find optimal algorithms whose combinatory complexity is proportional to the "cardinality" of $\mathfrak{N}$.

From (3.4) and (3.5) we get

$$(3.6)\quad \mathrm{comp}(\mathfrak{N},S,\varepsilon) \geq \mathrm{comp}(\mathfrak{N}) + m(\mathfrak{N}).$$

Furthermore if there exists $\varphi \in \Phi(\varepsilon)$ such that $\mathrm{comp}(\varphi(\mathfrak{N}(f))) \ll \mathrm{comp}(\mathfrak{N})$ for all $f \in \mathfrak{F}_0$ then

$$(3.7)\quad \mathrm{comp}(\mathfrak{N},S,\varepsilon) \cong \mathrm{comp}(\mathfrak{N})$$

Equations (3.6) and (3.7) motivate our interest in $\mathrm{comp}(\mathfrak{N})$.

Suppose that ε tends to zero. Then if $r(\mathfrak{N},S) > 0$, the fixed information $\mathfrak{N}$ is weak for the problem S for sufficiently small ε. However in many cases we can choose a permissible information operator $\mathfrak{N} = \mathfrak{N}(\varepsilon)$ such that $r(\mathfrak{N}(\varepsilon),S) < \varepsilon$ and solve our problem using the information $\mathfrak{N}(\varepsilon)$.

Let Ψ be a class of permissible information operators such that

$$(3.8) \quad \inf_{\mathfrak{N} \in \Psi} r(\mathfrak{N}, S) = 0.$$

Note that (3.8) means that we can solve the problem S for any ϵ using a suitably chosen information operator from Ψ.

Definition 3.2

We shall say $\text{comp}(\Psi, S, \epsilon)$ is the ϵ-complexity of the problem S in the class Ψ iff

$$(3.9) \quad \text{comp}(\Psi, S, \epsilon) = \inf_{\mathfrak{N} \in \Psi} \text{comp}(\mathfrak{N}, S, \epsilon). \qquad \blacksquare$$

Note that $\text{comp}(\Psi, S, \epsilon)$ is a nonincreasing function of ϵ. We shall prove in Section 9 that $\text{comp}(\Psi, S, \epsilon)$ can be an essentially arbitrary nonincreasing function of ϵ.

Our setting is sufficiently general that it includes problems for which information operators do not play a role. Examples are combinatorial problems and such problems of linear algebra as matrix multiplication and the direct solution of linear systems. For such problems the information operator is the identity operator $\mathfrak{N}(f) = f$ where f belongs to a finite dimensional space. Furthermore the information complexity $\text{comp}(I) = 0$ since there is no cost in computing $\mathfrak{N}(f) = f$. Note that $r(I, S) = 0$ for any problem S because $\mathfrak{N} = I$ is one to one. Thus we can define $\Psi = \{I\}$. Typically we seek the exact solution $\alpha = S(f)$; thus $\epsilon = 0$. The complexity is given by

$$\text{comp}(I, S, 0) = \inf_{\varphi:\ e(\varphi)=0} \text{comp}(\varphi(f)).$$

Therefore in algebraic and combinatorial complexity we seek an algorithm which finds $\alpha = S(f)$ and has minimal combinatory complexity.

Example 3.1. Matrix Multiplication

Let $f = [A,B]$ where A and B are nxn matrices. Then the matrix multiplication problem may be formulated as $S(f) = A*B$. Let $\mathfrak{N}(f) = f = [A,B]$. This means that all coefficients of A and B are known and we seek an algorithm with minimal combinatory complexity which yields the matrix $x(f) = A*B$. If the cost of each arithmetic operation is taken as unity, then

$$c_1 n^2 \leq comp(I,S,0) \leq c_2 n^{\beta}, \quad \beta = \log_2 7$$

for some positive constant c_1 and c_2. The actual value of comp(I,S,0) is unknown. ■

Example 3.2. Sorting

Let $f = [f_1,f_2,\ldots,f_n]$ where $f_i \in D$ and D is an ordered set. Define

$$S(f) = [f_{i_1},f_{i_2},\ldots,f_{i_n}]$$

where $f_{i_1} \leq f_{i_2} \leq \ldots \leq f_{i_n}$ and $i_1,\ldots,i_n$ is a permutation of $1,\ldots,n$. Then the sorting problem may be formulated as $\alpha = S(f)$. Let $\mathfrak{N}(f) = f$ and $\epsilon = 0$. We seek an algorithm which finds S(f) with minimal cost where the cost is taken as the number of comparisons. The complexity of this problem satisfies

$$comp(I,S,0) = (n \log_2 n)(1+o(1))$$ ■

Recently there has been an interest in finding ϵ-approximations to the solutions of algebraic and combinatory problems. For some problems the complexity $comp(I,S,\epsilon)$ is significantly smaller for positive ϵ than comp(I,S,0). Examples may be found in Garey and Johnson [76].

Example 3.3. Polynomial Zero

Let $f = [a_0, a_1, \ldots, a_n]$ where the a_i are the coefficients of an nth degree polynomial, $P(x)$. Let $P(\alpha) = 0$, $\alpha = S(f)$. For algorithms which require knowledge of the a_i, $\mathfrak{N}_1(f) = f$. On the other hand, there exist iterative algorithms requiring only that P and P' can be evaluated at any point and $\mathfrak{N}_2(f) = [x, P(x), p'(x)]$. ■

CHAPTER II

LINEAR THEORY

We assume that the solution operator S, the "restriction" operator T and the information operator N are linear.

4. CARDINALITY OF LINEAR INFORMATION

Let $N: F_1 \to F_3$ be a linear information operator where F_3 is a linear space. Let $\ker N = \{f: N(f) = 0\}$ be the kernel of N. We shall prove in Section 5 that the dependence of the diameter of information on N is only through the kernel of N. This suggests we should not distinguish between two information operators with the same kernel.

Let $N_1: F_1 \to F_3$ and $N_2: F_1 \to F_3'$ be two information operators where F_3' is a linear space not necessarily equal to F_3.

Definition 4.1

We shall say N_1 is contained in N_2 (briefly $N_1 \subset N_2$) iff $\ker N_2 \subset \ker N_1$.

We shall say N_1 is equivalent to N_2 (briefly $N_1 \asymp N_2$) iff $\ker N_1 = \ker N_2$. ■

Note that "$\asymp$" is an equivalence relation.

We want to show that $N_1 \subset N_2$ can be characterized by the rank of a certain matrix. We first briefly recall some facts on linear spaces. Let A be a linear subspace of F_1. Then there exists a linear subspace $A^{\perp}$ of F_1 such that

$$F_1 = A \oplus A^{\perp}. \tag{4.1}$$

In general, $A^{\perp}$ is not uniquely defined. However, if $\mathfrak{I}_1$ is a Hilbert space and A is closed, then there exists a unique orthogonal $A^{\perp}$ to A such that (4.1) holds. In either case $A^{\perp}$ is isomorphic to the quotient space $\mathfrak{I}_1/A$ and

(4.2) $\operatorname{codim} A \stackrel{\text{df}}{=} \dim A^{\perp} = \dim \mathfrak{I}_1/A.$

The space $A^{\perp}$ is called an algebraic complement of A in the space $\mathfrak{I}_1$.

Let $L_1, L_2, \ldots, L_m$ be linearly independent functionals. By

(4.3) $\mathfrak{N}_1 = [L_1, L_2, \ldots, L_m]^t$

we mean $\mathfrak{N}_1(f) = [L_1(f), L_2(f), \ldots, L_m(f)]^t \in \mathbb{C}^m$ where "t" denotes the transpose of a vector and $f \in \mathfrak{I}_1$.

Lemma 4.1

Let $\mathfrak{N}_1 = [L_1, L_2, \ldots, L_n]^t$ and $\mathfrak{N}_2 = [L_{n+1}, L_{n+2}, \ldots, L_{n+k}]^t$ be information operators.

(i) $\mathfrak{N}_1 \subset \mathfrak{N}_2$ iff $k \geq n$ and there exists a $n \times k$ matrix M such that

$$\mathfrak{N}_1 = M\mathfrak{N}_2 \text{ and rank } M = n.$$

(ii) Let $k = n$. Then $\mathfrak{N}_1 \subset \mathfrak{N}_2$ iff $\mathfrak{N}_1 \asymp \mathfrak{N}_2$. ■

Proof

(i) Suppose that $\ker \mathfrak{N}_2 \subset \ker \mathfrak{N}_1$. Let $\mathfrak{I}_1 = \ker \mathfrak{N}_2 \oplus (\ker \mathfrak{N}_2)^{\perp}$ and $(\ker \mathfrak{N}_2)^{\perp} = \operatorname{lin}(\xi_1, \xi_2, \ldots, \xi_k)$ where $L_{n+j}(\xi_i) = \delta_{ij}$ and δ_{ij} denotes the Kronecker delta. Then $f = f_0 + \sum_{i=1}^{k} L_{n+i}(f)\xi_i$ where $f_0 \in \ker \mathfrak{N}_2$ and $f \in \mathfrak{I}_1$. Since $f_0 \in \ker \mathfrak{N}_1$, we get

$$(4.4)\quad L_j(f) = \sum_{i=1}^{k} L_{n+i}(f)L_j(\xi_i) \quad \text{for } j = 1,2,\ldots,n.$$

This yields $\mathfrak{N}_1 = M\mathfrak{N}_2$ with $M = (L_j(\xi_i))$. Let $(\ker \mathfrak{N}_1)^{\perp} = \text{lin}\,(\eta_1,\eta_2,\ldots,\eta_n)$ where $L_j(\eta_i) = \delta_{ij}$. Set $f = \eta_i$ in (4.4) for $i = 1,2,\ldots,n$. Then $I = M[\mathfrak{N}_2(\eta_1),\mathfrak{N}_2(\eta_2),\ldots,\mathfrak{N}_2(\eta_n)]$ where I is the $n\times n$ identity matrix. This implies that rank $M = n$ and completes this part of the proof. Suppose now that $\mathfrak{N}_1 = M\mathfrak{N}_2$. Then $h \in \ker \mathfrak{N}_2$ implies $\mathfrak{N}_1(h) = M\mathfrak{N}_2(h) = 0$ which yields $\mathfrak{N}_1 \subset \mathfrak{N}_2$.

(ii) Suppose that $\mathfrak{N}_1 \subset \mathfrak{N}_2$. Due to the first part of Lemma 4.1 we get $\mathfrak{N}_1 = M\mathfrak{N}_2$ where the $n\times n$ matrix M is nonsingular. Then $\mathfrak{N}_2 = M^{-1}\mathfrak{N}_1$ which implies $\mathfrak{N}_2 \subset \mathfrak{N}_1$ and $\mathfrak{N}_1 \asymp \mathfrak{N}_2$. The second part of (ii) is trivial. Hence Lemma 4.2 is proven. ■

We now show that any information operator $\mathfrak{N}$ where $n = \text{codim ker } \mathfrak{N}$ is finite may be represented by n linearly independent linear functionals.

Lemma 4.2

Let $\mathfrak{N}$ be an information operator and $n = \text{codim ker } \mathfrak{N} < +\infty$. Then there exist linearly independent linear functionals $L_1,L_2,\ldots,L_n$ such that

$$\mathfrak{N} \asymp \mathfrak{N}_1 \quad \text{where} \quad \mathfrak{N}_1 = [L_1,L_2,\ldots,L_n]^t. \qquad ■$$

Proof

Let $(\ker \mathfrak{N})^{\perp} = \underset{n}{\text{lin}}(\xi_1,\xi_2,\ldots,\xi_n)$. Every element f has a unique representation $f = f_0 + \sum_{i=1}^{n} L_i(f)\xi_i$ where $f_0 \in \ker \mathfrak{N}$ and $L_1,L_2,\ldots,L_n$ are linearly independent linear functionals. Since $\ker \mathfrak{N} = \{f: L_i(f) = 0, i = 1,2,\ldots,n\} = \ker \mathfrak{N}_1$ we get $\mathfrak{N} \asymp \mathfrak{N}_1$. ■

Lemma 4.2 states that an information operator where n = codim ker $\mathfrak{N}$ is finite is equivalent to an information operator generated by n linearly independent linear functionals. Observe that to know $\mathfrak{N}_1(f)$ one has to evaluate n linear functionals. This suggests the following definition of the cardinality of $\mathfrak{N}$.

Definition 4.2

We shall say that card($\mathfrak{N}$) is the cardinality of the information $\mathfrak{N}$ iff

(4.5) card($\mathfrak{N}$) = codim ker $\mathfrak{N}$. ■

We shall prove in Section 5 that unless the cardinality of the information $\mathfrak{N}$ is sufficiently large, the diameter of information is infinity and the problem cannot be solved with this information.

To illustrate the concept of cardinality we consider two examples.

Example 4.1

Let $\mathfrak{N} = [L_1, L_2, \ldots, L_n]^t$. From the above considerations easily follows that card($\mathfrak{N}$) $\leq$ n and card($\mathfrak{N}$) = n iff $L_1, L_2, \ldots, L_n$ are linearly independent. ■

Example 4.2

Let $f: D \subset \mathbb{C}^m \to \mathbb{C}^m$ be a k times differentiable function. Let

$$\mathfrak{N}(f) = f^{(k)}(x) \quad \text{for } x \in D.$$

Note that $f(x) = [f_1(x), f_2(x), \ldots, f_m(x)]^t$ where $f_j: D \to \mathbb{C}^1$ is a scalar function. $f_j^{(k)}(x)$ can be represented by $\binom{m+k-1}{k}$ linearly independent functionals of the form $L(f) = \dfrac{\partial^k f}{\partial x_1^{p_1} \ldots \partial x_m^{p_m}}$ where $x = [x_1, x_2, \ldots, x_m]^t$ and $p_i \geq 0$, $p_1 + p_2 + \ldots + p_m = k$. This yields

$$\text{card}(\mathfrak{N}) = m\binom{m+k-1}{k} .$$

If $\mathfrak{N}(f) = [f(x), f'(x), \ldots, f^{(n-1)}(x)]$ is the so-called standard information, then

$$\text{card}(\mathfrak{N}) = \sum_{k=0}^{n-1} m\binom{m+k-1}{k} = m\binom{m+n-1}{n-1} .$$

This shows the dependence of cardinality on the dimension of the space $\mathbb{C}^m$. ■

We end this section by showing that there is a one-to-one correspondence between information operators and subspaces of $\mathfrak{F}_1$.

Let $\mathfrak{N}$ be an information operator with $\text{card}(\mathfrak{N}) = n$. Then $A(\mathfrak{N}) \stackrel{\text{df}}{=} \ker \mathfrak{N}$ has codimensionality equal to n. Furthermore $\mathfrak{N}_1 \asymp \mathfrak{N}_2$ implies $A(\mathfrak{N}_1) = A(\mathfrak{N}_2)$. We now show that the converse statement is also true.

Lemma 4.3

Let A be an arbitrary linear subspace of $\mathfrak{F}_1$ such that codim A = n. Then there exists a unique (up to the equivalence relation) information operator $\mathfrak{N}$ with $\text{card}(\mathfrak{N}) = n$ such that $A = \ker \mathfrak{N}$. ■

Proof

Let $\mathfrak{F}_1 = A \oplus A^{\perp}$ where $A^{\perp} = \text{lin}(\xi_1, \xi_2, \ldots, \xi_n)$. Then $f = f_0 + \sum_{i=1}^{n} L_i(f)\xi_i$ where $f_0 \in A$ and $L_i(\xi_j) = \delta_{ij}$. Define

$$\mathfrak{N} = [L_1, L_2, \ldots, L_n]^t .$$

Since $L_1, L_2, \ldots, L_n$ are linearly independent, $\text{card}(\mathfrak{N}) = n$ and $\ker \mathfrak{N} = A$. To prove the uniqueness, observe that if $A = \ker \mathfrak{N}_1 = \ker \mathfrak{N}_2$ then $\mathfrak{N}_1 \asymp \mathfrak{N}_2$. This completes the proof. ■

5. INDEX OF A LINEAR PROBLEM

We consider in this section linear information operators for the solution element $\alpha = S(f)$ where $S: \mathcal{F}_1 \to \mathcal{F}_2$ is a linear operator and $\mathcal{F}_0$ is defined as

$$(5.1) \quad \mathcal{F}_0 = \{f \in \mathcal{F}_1 : \|Tf\| \le 1\}$$

where $T: \mathcal{F}_1 \to \mathcal{F}_4 = T(\mathcal{F}_1)$ is a linear operator and $\mathcal{F}_4$ is a linear normed space over the real or complex field. We shall call T the restriction operator. This means we want to find an ϵ-approximation to the solution $\alpha = S(f)$ for all f such that $\|Tf\| \le 1$.

To stress the dependence on T we shall replace S by (S,T) in all basic definitions. For instance we shall refer to the problem (S,T), the diameter $d(\mathcal{N},S,T)$, etc., where $\mathcal{N}$ is a linear information operator.

Without loss of generality we choose a bound $\|Tf\| \le 1$ instead of $\|Tf\| \le c$ for a positive constant c. Indeed, let $T_1 = \frac{1}{c}T$. Then $\|T_1 f\| \le 1$ is equivalent to $\|Tf\| \le c$. It is easy to observe that $d(\mathcal{N},S,T) = cd(\mathcal{N},S,T_1)$ and all estimates on complexity are linear in c.

We now show that the dependence of $d(\mathcal{N},S,T)$ on $\mathcal{N}$ is only through the kernel of $\mathcal{N}$.

Lemma 5.1

$$(5.2) \quad d(\mathcal{N},S,T) = 2 \sup_{h \in V(0)} \|Sh\|$$

where $V(0) = \ker \mathcal{N} \cap \mathcal{F}_0$. (See (2.4).) ■

Proof

Set $c = 2 \sup_{h \in V(0)} \|Sh\|$. Let $f \in \mathcal{I}_0$ and $\tilde{f} \in V(f)$. Then $h = \frac{1}{2}(\tilde{f}-f) \in \ker \mathcal{N}$ and $\|Th\| \le 1$. This yields

$$\|S\tilde{f}-Sf\| = 2\|Sh\| \le c.$$

Taking the supremum with respect to f and $\tilde{f}$ we get $d(\mathcal{N},S,T) \le c$. To prove the reverse inequality, let $h \in V(0)$. Set $f = h$ and $\tilde{f} = -h$. Then $\tilde{f} \in V(f) = V(0)$ and

$$2\|Sh\| = \|S\tilde{f}-Sf\| \le d(\mathcal{N},S,T).$$

Thus $c \le d(\mathcal{N},S,T)$ which completes the proof. ■

From Lemma 5.1 we immediately get the following Corollary.

Corollary 5.1

If $\mathcal{N}_1 \subset \mathcal{N}_2$ then $d(\mathcal{N}_2,S,T) \le d(\mathcal{N}_1,S,T)$,

If $\mathcal{N}_1 \asymp \mathcal{N}_2$ then $d(\mathcal{N}_2,S,T) = d(\mathcal{N}_1,S,T)$. ■

In Section 2 we showed that the radius $r(\mathcal{N},S,T)$ is the intrinsic error of the information $\mathcal{N}$ and the problem (S,T). Due to Lemma 5.1 we get

$$c \le r(\mathcal{N},S,T) \le 2c$$

where $c = \sup_{h \in V(0)} \|Sh\|$. We now show when $r(\mathcal{N},S,T) = c$.

Lemma 5.2

If for any $f \in \mathcal{I}_0$ there exists $h_0 \in \ker \mathcal{N}$ such that $Th_0 = Tf$ then

$$r(\mathcal{N},S,T) = \sup_{h \in V(0)} \|Sh\|. \quad ■$$

Proof

Let $a = S(f-h_0)$. Then for any $\tilde{f} = f + h \in V(f)$, $h \in \ker \mathfrak{N}$, we get $\| a-S\tilde{f}\| = \| Sz\|$ where $z = h_0 + h \in \ker \mathfrak{N}$. Since $Th_0 = Tf$, $\| Tz\| = \| T(z-h_0)+Tf\| = \| T\tilde{f}\| \leq 1$. From (2.10) and Lemma 5.1 we get

$$r(\mathfrak{N},S,T) \leq \sup_{f\in\mathfrak{F}_0} \sup_{\tilde{f}\in V(f)} \| a-S\tilde{f}\| = \sup_{z\in V(0)} \| Sz\| = \frac{1}{2}d(\mathfrak{N},S,T).$$

Since $r(\mathfrak{N},S,T) \geq \frac{1}{2}d(\mathfrak{N},S,T)$ for any $\mathfrak{N}$, S and T, Lemma 5.2 is proven. ■

We want to examine when the diameter $d(\mathfrak{N},S,T)$ is equal to infinity. (Of course, $d(\mathfrak{N},S,T) = +\infty$ implies $r(\mathfrak{N},S,T) = +\infty$.) We begin with

Theorem 5.1

If $\ker \mathfrak{N} \cap \ker T \not\subset \ker S$ then $d(\mathfrak{N},S,T) = +\infty$. ■

Proof

Let $h \in \ker \mathfrak{N} \cap \ker T$ and $h \notin \ker S$. Then $T(ch) = 0$, $\mathfrak{N}(ch) = 0$ for any constant c. Then $\| S(ch)\| = |c|\ \| Sh\| \to +\infty$ with $|c| \to +\infty$. Due to Lemma 5.1 we get $d(\mathfrak{N},S,T) = +\infty$. ■

Theorem 5.1 states that $\ker \mathfrak{N} \cap \ker T$ has to be contained in $\ker S$ for $d(\mathfrak{N},S,T)$ to be finite. We prove that $\ker \mathfrak{N} \cap \ker T \subset \ker S$ implies that the cardinality of $\mathfrak{N}$ is at least as large as the "problem index". Let

$$\begin{aligned} &\ker T = (\ker T \cap \ker S) \oplus A(T,S), \\ &A(T,S) = \mathrm{lin}(\xi_1^*,\xi_2^*,\ldots,\xi_{n^*}^*) \end{aligned} \tag{5.3}$$

where $A(T,S)$ is an algebraic complement of $\ker T \cap \ker S$ in the space $\ker T$ and $\xi_1^*,\ldots,\xi_{n^*}^*$ form a basis of $A(T,S)$, $n^* = n^*(T,S) \leq +\infty$.

Definition 5.1

We shall say that index(S,T) = dim A(T,S) is the index of the problem (S,T). We shall sometimes write index (S,T) = n^*. ■

Note that index(S,T) = dim(ker T) - dim(ker T ∩ ker S) whenever either dim(ker T) or dim(ker T ∩ ker S) is finite. We are ready to prove the main result of this section.

Theorem 5.2

If card($\mathfrak{N}$) < index(S,T) then d($\mathfrak{N}$,S,T) = + ∞. ■

Proof

We show that ker $\mathfrak{N}$ ∩ ker T ⊄ ker S. Define $f = \sum_{i=1}^{n^*} c_i \xi_i^* \in A(T,S)$. We want to find a nonzero vector $(c_1,c_2,\ldots,c_{n^*})$ such that $f \in \ker \mathfrak{N}$. From Lemma 4.2 it follows that there exists an information operator $\mathfrak{N}_1 = [L_1,L_2,\ldots,L_m]^t$ where m = card($\mathfrak{N}$) < index(S,T) such that ker $\mathfrak{N}_1$ = ker $\mathfrak{N}$. Thus $f \in \ker \mathfrak{N}$ iff $L_j(f) = \sum_{i=1}^{n^*} c_i L_j(\xi_i^*) = 0$ for j = 1,2,...,m. Hence we get a homogeneous system of m linear equations in n^* unknowns. Since $m < n^*$, there exists a nonzero vector $(c_1,c_2,\ldots,c_{n^*})$ which is a solution of the system. Thus $0 \neq f \in A(T,S) \cap \ker \mathfrak{N}$, This means that a nonzero f belongs to ker $\mathfrak{N}$ ∩ ker T and $f \notin \ker S$. Due to Theorem 5.1 we get d($\mathfrak{N}$,S,T) = + ∞. ■

Theorem 5.2 states that every information operator with cardinality less than the index of S and T does not supply enough information to solve the problem. For index(S,T) = + ∞ we get the following corollary.

Corollary 5.2

If index(S,T) = $+\infty$ then the problem (S,T) cannot be solved by any information operator with finite cardinality. ■

We illustrate the above results for some restriction operators T. We begin with T = 0.

Lemma 5.3 (No restriction operator)

Let T = 0. Then

(i) $d(\mathfrak{N},S,0)$ is either zero or infinity. More precisely $\ker \mathfrak{N} \not\subset \ker S$ implies $d(\mathfrak{N},S,0) = +\infty$, $\ker \mathfrak{N} \subset \ker S$ implies $d(\mathfrak{N},S,0) = 0$.

(ii) $\text{index}(S,0) = \dim(\ker S)^{\perp}$ is finite iff S is a finite dimensional operator, i.e., $\dim S(\mathfrak{F}_1) < +\infty$. ■

Proof

Since T = 0, $\mathfrak{F}_0 = \mathfrak{F}_1$ and $\ker T = \mathfrak{F}_1$. If $\ker \mathfrak{N} \cap \ker T = \ker \mathfrak{N} \not\subset \ker S$ then $d(\mathfrak{N},S,0) = +\infty$ due to Theorem 5.1. If $\ker \mathfrak{N} \subset \ker S$ then Sh = 0 for all $h \in V(0) = \ker \mathfrak{N} \subset \ker S$. Thus $d(\mathfrak{N},S,0) = 0$ by Lemma 5.2. This proves (i).

From (5.3) we get $A(0,S) = (\ker S)^{\perp}$ and $\text{index}(S,0) = \dim(\ker S)^{\perp}$. It is well known that index(S,0) is finite iff S is a finite dimensional operator. This proves Lemma 5.3. ■

As our second illustration consider $T = D^k$, $k \geq 0$, i.e., $Tf = f^{(k)}$ for a scalar function f. If S is a one-to-one operator then

$$A(T,S) = \ker T = \{f: \ f^{(k)} \equiv 0\}$$

and $\text{index}(S,D^k) = \dim(\ker T) = k$. Hence we have to compute k linear functionals to assure that $\text{card}(\mathfrak{N}) \geq \text{index}(S,D^k)$ and $\ker \mathfrak{N} \cap \ker T = \ker S = \{0\}$.

6. OPTIMAL LINEAR INFORMATION OPERATORS

Assume that $n^* = \text{index}(S,T) < +\infty$. We construct an information operator $\mathfrak{N}^*$ with $\text{card}(\mathfrak{N}^*) = \text{index}(S,T)$ such that $\ker \mathfrak{N}^* \cap \ker T \subset \ker S$. Recall that $A(T,S)$ is defined by (5.3) and $A(T,S) = \text{lin}(\xi_1^*, \xi_2^*, \ldots, \xi_{n^*}^*)$. Let $\mathfrak{I}_1 = A(T,S) \oplus A(T,S)^\perp$ and $f = \sum_{i=1}^{n^*} L_i^*(f)\xi_i^* + f_1$ where $f_1 \in A(T,S)^\perp$ and $L_i^*(\xi_j^*) = \delta_{ij}$. Then $\mathfrak{N}^* = A(T,S)^\perp$.

Lemma 6.1

Let

$$\mathfrak{N}^* = [L_1^*, L_2^*, \ldots, L_{n^*}^*]^t. \tag{6.1}$$

Then $\ker \mathfrak{N}^* \cap \ker T \subset \ker S$. ■

Proof

Let $f \in \ker T$. Then $f = f_0 + \sum_{i=1}^{n^*} L_i^*(f)\xi_i^*$ where $f_0 \in \ker T \cap \ker S$. If $f \in \ker \mathfrak{N}^*$ then $L_i^*(f) = 0$ for $i = 1,2,\ldots,n^*$ and $f = f_0 \in \ker S$. This proves Lemma 6.1. ■

To simplify further considerations and to assure that $\ker \mathfrak{N} \cap \ker T \subset \ker S$ we shall consider throughout this section only information operators $\mathfrak{N}$ such that $\mathfrak{N}^* \subset \mathfrak{N}$. (This means $\ker \mathfrak{N} \subset \ker \mathfrak{N}^*$ and $\ker \mathfrak{N} \cap \ker T \subset \ker \mathfrak{N}^* \cap \ker T \subset \ker S$ due to Lemma 6.1.)

We show the diameter of $d(\mathfrak{N},S,T)$ can be computed in terms of the inverse operator T^{-1} which is defined as follows. (T is not one-to-one in general.) Recall that $\mathfrak{I}_4 = T(\mathfrak{I}_1)$ and

$$\mathfrak{I}_1 = \ker T \cap \ker S \oplus A(T,S) \oplus (\ker T)^\perp. \tag{6.2}$$

Thus $f = f_0 + f_1 + f_2 \in \mathfrak{H}_1$ where $f_0 \in \ker T \cap \ker S$, $f_1 \in A(T,S)$ and $f_2 \in (\ker T)^{\perp}$. From Lemma 5.1 it follows that the dependendence of T on the diameter $d(\mathfrak{N},S,T)$ is only through the kernel of $\mathfrak{N}^*$. Let $f \in \ker \mathfrak{N}^*$. Then $f = f_0 + f_2$ since $f_1 \in A(T,S) = (\ker \mathfrak{N}^*)^{\perp}$. Define a linear operator $T^{-1}: \mathfrak{H}_4 \to \mathfrak{H}_1$ such that

$$(6.3) \quad T^{-1}z = f_2$$

where $z = Tf$.

We check that T^{-1} is well-defined. Let $z = Tf = Tg$ where $g = g_0 + g_2 \in \ker \mathfrak{N}^*$. Since $T(f-g) = 0$, $f-g = (f_0-g_0) + (f_2-g_2) \in \ker T$ which yields $f_2 = g_2$. This proves that $T^{-1}z$ does not depend on a particular choice of pre-image of z. Hence T^{-1} is well-defined.

As an example observe that $T = 0$ implies $\mathfrak{H}_4 = \{0\}$ and $0^{-1} = 0$.

Let $K: \mathfrak{H}_4 \to \mathfrak{H}_2$ be a linear operator and let B be a linear subspace of $\mathfrak{H}_4$. Denote

$$(6.4) \quad \|K\|_B \stackrel{df}{=} \sup_{\|z\| \le 1, z \in B} \|Kz\| .$$

We are ready to prove

Lemma 6.2

Let $\mathfrak{N}^* \subset \mathfrak{N}$. Then

$$(6.5) \quad d(\mathfrak{N},S,T) = 2\|ST^{-1}\|_{T(\ker \mathfrak{N})}. \qquad \blacksquare$$

Proof

Note that ker $\mathfrak{N} \subset$ ker $\mathfrak{N}^*$ assures that we can use T^{-1} defined by (6.3) for the linear subspace T(ker $\mathfrak{N}$). Lemma 5.1 states that $d(\mathfrak{N},S,T) = 2 \sup \|Sh\|$ where $\mathfrak{N}h = 0$ and $\|Th\| \leq 1$. Let $h = h_0 + h_2$ due to (6.2). Then $Sh = Sh_2$ and $Th = Th_2$. Let $z = Th \in T(\ker \mathfrak{N})$. Observe that $T^{-1}z = h_2$ and $ST^{-1}z = Sh_2$. This proves (6.5). ■

Lemma 6.2 states that the diameter $d(\mathfrak{N},S,T)$ is equal to twice the norm of the linear operator $K = ST^{-1}$ in a certain linear subspace $B = T(\ker \mathfrak{N})$. This suggests the following problem. For a fixed integer n find the most relevant information operator $\mathfrak{N}$, $\text{card}(\mathfrak{N}) \leq n$, that is the operator which minimizes $d(\mathfrak{N},S,T)$ among all information operators with cardinality $\leq n$. This is equivalent, as we shall prove, to finding a linear subspace B with codim $B \leq n-n^*$ which minimizes $\|K\|_B$ among all linear subspaces of codimension $\leq n-n^*$.

To formalize this problem let Ψ_n be the class of all information operators $\mathfrak{N}$ such that $\mathfrak{N}^* \subset \mathfrak{N}$, $\text{card}(\mathfrak{N}) \leq n$ where $n \geq \text{index}(S,T)$.

Definition 6.1

We shall say $d(n,S,T)$ is the n-th minimal diameter of information iff

$$(6.6) \qquad d(n,S,T) = \begin{cases} \inf_{\mathfrak{N} \in \Psi_n} d(\mathfrak{N},S,T) & \text{if } n \geq \text{index}(S,T) \\ +\infty & \text{if } n < \text{index}(S,T) \end{cases}$$

We shall say $\mathfrak{N}_n^{oi}$ is an n-th optimal information iff

$$(6.7) \qquad d(n,S,T) = d(\mathfrak{N}_n^{oi},S,T). \qquad ■$$

We define $d(n,S,T) = +\infty$ for $n <$ index(S,T) since for any $\mathfrak{N}$ with cardinality less than index(S,T), $d(\mathfrak{N},S,T) = +\infty$. See Theorem 5.2. We illustrate Definition 6.1 by the following example.

Example 6.1

Let dim $\mathfrak{I}_1 = +\infty$ and let $T = I$ be the identity operator. Define $S = cI$ for a positive constant c. Then

(6.8) $d(n,cI,I) = 2c, \quad \forall n.$

Indeed let $\mathfrak{N} \in \Psi_n$. Then $\ker \mathfrak{N} \neq \{0\}$ and $d(\mathfrak{N},S,T) = 2\|cI\|_{\ker \mathfrak{N}} = 2c$. Note that $\varphi(\mathfrak{N}(f)) \equiv 0$ is an optimal error algorithm since $e(\varphi) = c = \frac{1}{2}d(\mathfrak{N},S,T) = r(\mathfrak{N},S,T)$. This means that no matter how many linear functionals are computed the zero of the space $\mathfrak{I}_1$ is the best approximation to the solution $Sf = cf$ for some f such that $\|f\| \leq 1$. See Schultz [73] for related material.

However for the identity information operator $\mathfrak{N}(f) \equiv f$ we get $\ker \mathfrak{N} = \{0\}$ and $d(I,cI,I) = 0$. Note that $\text{card}(I) = +\infty$. This shows that $d(n,S,T)$ can be a discontinuous function of n at infinity. ■

From Example 6.1 we get the following corollary.

Corollary 6.1

For every ε (no matter how large) there exists a linear problem (S,T) with finite index for which one cannot find an ε-approximation using any finite number of linear functionals. ■

We show that the n-th minimal diameter and the n-th optimal information are fully determined by the operator $K = ST^{-1}$. Let

(6.9) $$b(m,K) = \inf_{B \subset \mathfrak{I}_4,\ \text{codim } B \leq m} 2\|K\|_B$$

be the m-th minimal norm of the linear operator K.

Suppose there exists a sequence $\{B_m\}$, $m \geq 0$, such that

(6.10) $b(m,K) = 2\|K\|_{B_m}$ and $\text{codim } B_m \leq m$.

Let $\mathfrak{Z}_4 = B_m \oplus B_m^{\perp}$ and

$$(6.11)\quad g = g_0 + \sum_{i=1}^{k} L_{im}(g)\ \eta_{im}$$

where $g_0 \in B_m$ and $B_m^{\perp} = \mathrm{lin}(\eta_{1m}, \eta_{2m}, \ldots, \eta_{k,m})$ with $k = k(m) = \mathrm{codim}\ B_m \leq m$. We shall call B_m an m-th minimal subspace of the linear operator K.

Recall that $L_1^*, L_2^*, \ldots, L_{n^*}^*$ form $\mathfrak{N}^*$. See (5.3) and (6.1). Define

$$(6.12)\quad \mathfrak{N}_n = [L_1^*, L_2^*, \ldots, L_{n^*}^*, L_{1,n-n^*T}, \ldots, L_{k(n-n^*),n-n^*T}]^t.$$

We are ready to prove the main result of this section.

Theorem 6.1

The information $\mathfrak{N}_n$ defined by (6.12) is the n-th optimal information and

$$(6.13)\quad d(\mathfrak{N}_n, S, T) = d(n, S, T) \asymp b(n-n^*, K),\ K = ST^{-1},\ n^* = \mathrm{index}(S,T). \qquad \blacksquare$$

Proof

To prove Theorem 6.1 we need two lemmas.

Lemma 6.3

Let B be any linear subspace of $\mathfrak{Z}_4$ with $\mathrm{codim}\ B = k < +\infty$. Then there exists a unique information $\mathfrak{N}$ such that

(i) $\mathfrak{N}^* \subset \mathfrak{N}$,

(ii) $T(\ker \mathfrak{N}) = B$,

(iii) $\mathrm{card}(\mathfrak{N}) = k + n^*$. $\blacksquare$

Proof of Lemma 6.3

Let $\mathfrak{Z}_4 = B \oplus B^{\perp}$ and $B^{\perp} = \mathrm{lin}(\eta_1, \eta_2, \ldots, \eta_k)$. Thus for every $g \in \mathfrak{Z}_4$ we have $g = g_0 + \sum_{i=1}^{k} L_i(g)\eta_i$ where $g_0 \in B$ and $L_i(\eta_j) = \delta_{ij}$. Define

(6.14) $\mathfrak{N} = [L_1^*, L_2^*, \ldots, L_{n^*}^*, L_1T, L_2T, \ldots, L_kT]^t$.

Then $\ker \mathfrak{N} = \{f: L_i^*(f) = 0, L_j(Tf) = 0, i = 1,2,\ldots,n^*, j = 1,2,\ldots,k\} \subset \ker \mathfrak{N}^*$. This proves (i). Let $h \in \ker \mathfrak{N}$. Then $L_i(Th) = 0$ for $i = 1,2,\ldots,k$ and $Th \in B$. Hence $T(\ker \mathfrak{N}) \subset B$. Now let g be an arbitrary element of B, i.e., $L_i(g) = 0$ for $i = 1,2,\ldots,k$. Since $g \in \mathfrak{I}_4$, there exists $f \in \mathfrak{I}_1$ such that $g = Tf$. Decompose $f = f_0 + f_1 + f_2$ where $f_0 \in \ker T \cap \ker S$, $f_1 \in A(T,S)$ and $f_2 \in (\ker T)^{\perp}$. See (6.2). Then $g = Tf = Tf_2$. Since $L_i^*(f_2) = 0$ for $i = 1,2,\ldots,n^*$ and $L_i(Tf_2) = 0$ for $i = 1,2,\ldots,k$ we get $f_2 \in \ker \mathfrak{N}$ and $g = Tf_2 \in T(\ker \mathfrak{N})$. This yields $T(\ker \mathfrak{N}) = B$ which proves (ii).

To prove that $\text{card}(\mathfrak{N}) = k+n^*$ we show that $L_1^*, \ldots, L_{n^*}^*, L_1T, \ldots, L_kT$ are linearly independent. See Example 4.1. Assume that

$$\left(\sum_{i=1}^{n^*} c_i L_i^* + \sum_{i=1}^{k} d_i L_i T \right) f = 0, \quad \forall f \in \mathfrak{I}_1.$$

Set $f = \xi_i^*$ where $\xi_1^*, \xi_2^*, \ldots, \xi_{n^*}^*$ form a basis of $A(T,S)$. See (5.3). Then $T\xi_i^* = 0$ and $L_j^*(\xi_i^*) = \delta_{ij}$. This yields $c_i = 0$ for $i = 1,2,\ldots,n^*$. Now let $\eta_i = Tf_i$ and set $f = f_i$. Since $L_j(Tf_i) = L_j(\eta_i) = \delta_{ij}$, we get $d_i = 0$ for $i = 1,2,\ldots,k$. This proves that $\text{card}(\mathfrak{N}) = k+n^*$.

We now show the uniqueness of $\mathfrak{N}$. Suppose that an information operator $\mathfrak{N}_1 = [\tilde{L}_1, \tilde{L}_2, \ldots, \tilde{L}_{k+n^*}]^t$ satisfies (i)-(iii). Thus $\ker \mathfrak{N}_1 \subset \ker \mathfrak{N}^*$ means that $h \in \ker \mathfrak{N}_1$ implies $L_i^*(h) = 0$ for $i = 1,2,\ldots,n^*$. Next $T(\ker \mathfrak{N}_1) = B$ means that $h \in \ker \mathfrak{N}_1$ implies $Th \in B$, i.e., $L_i(Th) = 0$, $i = 1,2,\ldots,k$. Thus $h \in \ker \mathfrak{N}$ and $\ker \mathfrak{N}_1 \subset \ker \mathfrak{N}$. Since $\text{card}(\mathfrak{N}_1) = \text{card}(\mathfrak{N})$, from Lemma 4.1 we get $\mathfrak{N}_1 \asymp \mathfrak{N}$. This completes the proof of Lemma 6.3. ■

Due to the uniqueness of $\mathfrak{N}$ we shall write $\mathfrak{N} = T^{-1}(B)$. Note that $\mathfrak{N}_n$ defined by (6.12) is equal to $T^{-1}(B_{n-n^*})$ where B_{n-n^*} is the $(n-n^*)$th minimal subspace of K.

Lemma 6.4

Let card($\mathfrak{N}$) = n. Then codim T(ker $\mathfrak{N}$) $\le$ n-n*. ■

Proof

Let B = T(ker $\mathfrak{N}$) and let k = codim B. From Lemma 6.3, $\mathfrak{N}_1 = T^{-1}(B)$ has the properties $\mathfrak{N}^* \subset \mathfrak{N}_1$ and card($\mathfrak{N}_1$) = k+n*. Repeating a part of the proof of Lemma 6.3, it is easy to show that ker $\mathfrak{N} \subset$ ker $\mathfrak{N}_1$. From Lemma 4.1 we get $\mathfrak{N}_1 = M\mathfrak{N}$ where the (k+n*) $\times$ n matrix M has rank M = k+n*. This is possible only if k $\le$ n-n* which completes the proof. ■

We proceed to prove Theorem 6.1. From Lemma 6.3 we know

$$d(\mathfrak{N}_n, S, T) = 2 \| K \|_{B_{n-n^*}} = b(n-n^*, K).$$

Let $\mathfrak{N}$ be any information operator from Ψ_n. From Lemma 6.4 we get dim T(ker $\mathfrak{N}$) $\le$ n-n* and

$$d(\mathfrak{N}, S, T) = 2 \| K \|_{T(\ker \mathfrak{N})} \ge b(n-n^*, K) = d(\mathfrak{N}_n, S, T).$$

This proves that $\mathfrak{N}_n$ *is the n-th optimal information and* $d(\mathfrak{N}, S, T) = d(n, S, T) = b(n-n^*, K)$. This completes the proof of Theorem 6.1. ■

If d(n,S,T) $\ge 2\varepsilon$ then it is impossible to find an ε-approximation no matter which information operator $\mathfrak{N}$ with card($\mathfrak{N}$) $\le$ n is used. In this case we have to increase n and possibly find such m > n that d(m,S,T) $< 2\varepsilon$. This motivates our interest in the dependence of the n-th optimal diameter d(n,S,T) on n. Note that d(n,S,T) is a nonincreasing function of n.

Definition 6.2

We shall say d(S,T) is the <u>diameter of problem error</u> in the class of information of finite cardinality iff

(6.15) $d(S,T) = \lim_{n\to\infty} d(n,S,T)$.

We shall say that the problem (S,T) is strongly non-computable *if* $d(S,T) = +\infty$, is ε-non-computable if $d(S,T) \geq 2\varepsilon > 0$ and is convergent if $d(S,T) = 0$. ■

We now show that the diameter of problem error $d(S,T)$ can be any number. This shows that for any ε there exist linear problems which are ε-non-computable.

Lemma 6.5

Let $\delta \in [0, +\infty]$. Then there exists a linear problem (S,T) such that

(6.16) $d(S,T) = \delta$. ■

Proof

Let $\delta = +\infty$. Define $T = 0$ and let S be a one-to-one operator. From Lemma 5.3 we get $\text{index}(S,0) = +\infty$ for infinite dimensional $\mathfrak{F}_1$. Thus by Theorem 5.2, $d(n,S,0) = +\infty$ for any finite n and $d(S,T) = +\infty = \delta$. Now let $\delta \in [0, +\infty)$. From Example 6.1 we get $d(n,\frac{\delta}{2}I,I) = \delta$ for any n. Thus $d(S,T) = \delta$ which completes the proof. ■

In the next section we show when the problem (S,T) is convergent and how to find the n-th minimal information.

7. CONVERGENCE AND MINIMAL SUBSPACES FOR A HILBERT SPACE

In this section we assume that $\mathfrak{F}_4$ is a Hilbert space. Recall that the problem (S,T) is convergent if $d(n,S,T) \to 0$ when $n \to +\infty$. See (6.8).

Theorem 7.1

Let $\mathfrak{F}_4$ be a Hilbert space and $\text{index}(S,T) < +\infty$. The problem (S,T) is convergent iff $K = ST^{-1}$ is compact. ■

Proof

Recall that $K: \mathfrak{F}_4 \to \mathfrak{F}_2$ is compact iff K may be uniformly approximated by finite dimensional linear operators, i.e., there exists $\{K_n\}$ such that

(i) $K_n: \mathfrak{F}_4 \to \mathfrak{F}_2$, K_n is linear,

(ii) $\dim K_n(\mathfrak{F}_4) < +\infty$

(iii) $\lim_{n\to\infty} \| K - K_n \| = 0.$

Suppose that the problem (S,T) is convergent, i.e., $\lim_{n\to\infty} d(n,S,T) = 0$. This means there exists a sequence of information operators $\{\mathfrak{N}_n\}$ such that $\text{card}(\mathfrak{N}_n) \le n$ and $\lim_{n\to\infty} d(\mathfrak{N}_n,S,T) = 0$. Due to Lemma 6.2 and Lemma 6.4 we get

$$d(\mathfrak{N}_n,S,T) = 2\| K\|_{B_n} \to 0, \quad B_n = T(\ker \mathfrak{N}_n)$$

and $\text{codim } B_n \le n\text{-}n^*$. Let $\mathfrak{F}_4 = B_n \oplus B_n^{\perp}$ and let $g = g_0 + g_1$ where $g_0 \in B_n$, $g_1 \in B_n^{\perp}$ and $(g_0,g_1) = 0$. Define

(7.1) $K_n g = Kg_1$.

Then K_n is a linear operator from $\mathfrak{F}_4$ to $\mathfrak{F}_2$ and $\dim K_n(\mathfrak{F}_4) = \dim K(B_n^{\perp}) \le \dim B_n^{\perp} \le n\text{-}n^* < +\infty$. Furthermore

$$\| (K-K_n)g \| = \| Kg_0 \| \le \| K \|_{B_n} \| g_0 \| = \| K \|_{B_n} \sqrt{\| g \|^2 - \| g_1 \|^2} \le$$

$$\le \| K \|_{B_n} \| g \|.$$

This proves that $\| K-K_n \| \le \| K \|_{B_n} \to 0$ which means that K is compact. This completes the first part of the proof.

Suppose now that K is compact. Then K may be uniformly approximated by $\{K_n\}$. Define $B_p = \ker K_n$ where $p = \operatorname{codim} \ker K_n < +\infty$. From Lemma 6.3 we know there exists a unique information operator $\mathfrak{N}_{p+n^*} = T^{-1}(B_p)$ such that $\mathfrak{N}^* \subset \mathfrak{N}_{p+n^*}$ and $\operatorname{card}(\mathfrak{N}_{p+n^*}) = p+n^*$. From Lemma 6.2 we get $d(\mathfrak{N}_{p+n^*},S,T) = 2\| K \|_{B_p}$. Since $Kg = (K-K_n)g$ for any $g \in B_p$, we have $\| Kg \| \le \| K-K_n \| \| g \|$. Hence $\| K \|_{B_p} \le \| K-K_n \| \to 0$ with $n \to +\infty$. Finally

$$d(p+n^*,S,T) \le d(\mathfrak{N}_{p+n^*},S,T) \le 2\| K \|_{B_p} \to 0.$$

Since $d(n,S,T)$ is a nonincreasing sequence, then $\lim_n d(n,S,T) = 0$ which completes the proof. ■

Remark 7.1

It is possible to generalize Theorem 7.1 for the case that $\mathfrak{Z}_4$ is not a Hilbert space and for problems (S,T) with nonzero d(S,T). We shall not pursue the generalization in this paper. ■

Theorem 7.1 states necessary and sufficient conditions for the problem (S,T) to be convergent. Note that in many cases $K = ST^{-1}$ is not compact. This holds, for instance, for $S = T = I$ and infinite dimensional $\mathfrak{Z}_4$.

We show how to find minimal subspaces of K assuming that K is compact and $K(\mathfrak{Z}_4) \subset \mathfrak{Z}_4$. Let K^* be an adjoint operator of K. Define a self-adjoint and compact operator

(7.2) $K_1 \stackrel{df}{=} K^*K: \mathcal{F}_4 \to \mathcal{F}_4$.

Decompose $\mathcal{F}_1 = \ker K_1 \oplus (\ker K_1)^\perp$ where the orthogonal complement $(\ker K)^\perp$ is spanned by eigenvectors of K_1, i.e.,

(7.3) $(\ker K_1)^\perp = \mathrm{lin}(\xi_1, \xi_2, \ldots, \xi_r)$, $r \leq +\infty$,

$K_1\xi_i = \lambda_i\xi_i$

where $\lambda_i > 0$ and $\lambda_1 \geq \lambda_2 \geq \ldots$, $(\xi_i, \xi_j) = \delta_{ij}$. If r is finite we formally put $\lambda_i = 0$ and $\xi_i = 0$ for $i \geq r+1$. Due to compactness of K, $\lim_i \lambda_i = 0$. Every element of $f \in \mathcal{F}_4$ has the unique decomposition $f = f_0 + \sum_{i=1}^{\infty} (f, \xi_i)\xi_i$ where $f_0 \in \ker K_1$. Define

(7.4) $B_{n-n^*} = \ker K_1 \oplus \mathrm{lin}(\xi_{n-n^*+1}, \ldots, \xi_r)$, $n \geq n^*$,

and an information operator

(7.5) $\mathfrak{N}_n(f) = [L_1^*(f), \ldots, L_{n^*}^*(f), (Tf, \xi_1), \ldots, (Tf, \xi_{n-n^*})]^t$

where $L_1^*, \ldots L_{n^*}^*$ are given by (6.1). We are ready to prove

Theorem 7.2

Let $\mathcal{F}_4$ be a Hilbert space, $n^* = \mathrm{index}(S,T) < +\infty$, and let $K = ST^{-1}$ be a compact operator such that $K(\mathcal{F}_4) \subset \mathcal{F}_4$.

The information operator $\mathfrak{N}_n$ defined by (7.5) is the n-th optimal information, B_{n-n^*} defined by (7.4) is the $(n-n^*)$th minimal subspace of K and

(7.6) $d(\mathfrak{N}_n, S, T) = d(n, S, T) = b(n-n^*, K) = 2\sqrt{\lambda_{n-n^*+1}}$. ■

Proof

We first show that B_k is the k-th minimal subspace, $k = n-n^*$. Let $f \in B_k$. Then $f = f_0 + \sum_{i=k+1}^{\infty} (f,\xi_i)\xi_i$ where $f_0 \in \ker K_1$ and $\|Kf\|^2 = (Kf,Kf) = (K_1 f,f) = \sum_{i=k+1}^{\infty} |(f,\xi_i)|^2 \lambda_i \le \lambda_{k+1}\|f\|^2$. Since this bound is sharp we get $\|K\|_{B_k} = \sqrt{\lambda_{k+1}}$. Let now B be any linear subspace such that codim $B \le k$. Then $\mathcal{F} = B \oplus B^{\perp}$ and $B^{\perp} = \mathrm{lin}(\eta_1,\eta_2,\ldots,\eta_m)$ where $m = \mathrm{codim}\, B \le k$. Furthermore $f = f_0 + \sum_{i=1}^{m} L_i(f)\eta_i$ for certain linear functionals $L_1,L_2,\ldots,L_m$ and $f_0 \in B$. Thus $f \in B$ iff $L_i(f) = 0$ for $i = 1,2,\ldots,m$. Let $f = \sum_{i=1}^{k+1} c_i\xi_i$. Then $L_i(f) = 0$ for $i = 1,2,\ldots,m$ is equivalent to $M\vec{c} = \vec{0}$ where $M = (L_i(\xi_j))$ is the $m \times (k+1)$ matrix and $\vec{c} = [c_1,c_2,\ldots,c_{k+1}]^t$. Since $m \le k+1$ there always exists a nonzero solution $\vec{c}$ and therefore a nonzero $f = \sum_{i=1}^{k+1} c_i\xi_i$ which belongs to B. Then

$$\|Kf\|^2 = (K_1 f,f) = \sum_{i=1}^{k+1} |c_i|^2 \lambda_i \ge \lambda_{k+1}\|f\|^2$$

which yields $\|K\|_B \ge \sqrt{\lambda_{k+1}} = \|K\|_{B_k}$. This proves that B_k is the k-th minimal subspace and $b(n-n^*,K) = 2\sqrt{\lambda_{n-n^*+1}}$.

Note that $B_k^{\perp} = \mathrm{lin}(\xi_1,\xi_2,\ldots,\xi_k)$ and $L_{ik}(f) = (f,\xi_i)$ for $i = 1,2,\ldots,k$. See (6.11). Thus the information operator $\mathfrak{N}_n$ defined by (7.5) is identical with (6.12). From Theorem 6.1 we get that $\mathfrak{N}_n$ is the n-th optimal information and

$$d(\mathfrak{N}_n,S,T) = d(n,S,T) = b(n-n^*,k) = 2\sqrt{\lambda_{n-n^*+1}}\,.$$

This completes the proof. ■

The information $\mathfrak{N}_n$ supplies the best possible information on the problem (S,T) in the class Ψ_n. Note that the evaluation of (Tf, ξ_i) means that we compute the ith component of Tf in the eigenvalue decomposition of K_1.

8. OPTIMAL ERROR ALGORITHMS FOR LINEAR CASE

In this section we give optimal error algorithms assuming that S, $\mathfrak{N}$ and T are linear operators.

Lemma 8.1

Suppose that for every $f \in \mathfrak{I}_0$ there exists $\tilde{f}$ such that

(8.1) $\mathfrak{N}(\tilde{f}) = \mathfrak{N}(f)$ and $T(\tilde{f}) = 0$.

Then the algorithm $\varphi(\mathfrak{N}(f)) = S\tilde{f}$ is an optimal error algorithm and

(8.2) $e(\varphi) = r(\mathfrak{N},S,T) = \frac{1}{2}\, d(\mathfrak{N},S,T)$. ■

Proof

Since $h = f - \tilde{f} \in \ker \mathfrak{N}$ and $\| Th \| \le 1$, we get from (5.2)

$$\| \varphi(\mathfrak{N}(f)) - S(f) \| = \| Sh \| \le \sup_{h \in \ker \mathfrak{N},\ \| Th \| \le 1} \| Sh \| = \frac{1}{2}\, d(\mathfrak{N},S,T).$$

Thus by (2.13) $e(\varphi) = \sup_f \| \varphi(\mathfrak{N}(f) - S(f) \| \le \frac{1}{2}\, d(\mathfrak{N},S,T)$. However (2.11) and (2.15) state that $e(\varphi) \ge r(\mathfrak{N},S,T) \ge \frac{1}{2}\, d(\mathfrak{N},S,T)$ which proves (8.2). ■

Let $\mathfrak{N} = [L_1, L_2, \ldots, L_n]^t$ and card $(\mathfrak{N}) = n$. To minimize combinatory complexity (see Section 9) we seek optimal (or close to optimal) error algorithms which are linear, i.e., $\varphi(\mathfrak{N}(f)) = \sum_{i=1}^{n} L_i(f) g_i$ for some elements $g_1, g_2, \ldots, g_n$.

We first consider the case that $k = \dim \mathfrak{I}_1 < +\infty$. Note that for every information operator $\mathfrak{N}$, $\text{card}(\mathfrak{N}) \le \dim \mathfrak{I}_1$.

Lemma 8.2

Let $\mathfrak{N} = [L_1, L_2, \ldots, L_k]^t$.

If card $\mathfrak{N} = k = \dim \mathfrak{F}_1 < +\infty$ then $d(\mathfrak{N}, S, T) = 0$ and there exist elements $g_1, g_2, \ldots, g_k$ such that

$$\varphi(\mathfrak{N}(f)) \stackrel{df}{=} \sum_{i=1}^{k} L_i(f) S g_i = Sf$$

is a linear interpolatory optimal error algorithm. ■

Proof

Since card $\mathfrak{N} = \dim \mathfrak{F}_1 < +\infty$, $\ker \mathfrak{N} = \{0\}$ and $(\ker \mathfrak{N})^{\perp} = \mathfrak{F}_1$. From Lemma 5.1 we get $d(\mathfrak{N}, S, T) = 0$. There exist $g_1, g_2, \ldots, g_k$ such that $\mathfrak{F}_1 = \operatorname{lin}(g_1, g_2, \ldots, g_k)$ and $L_i(g_j) = \delta_{ij}$. Then for every $f \in \mathfrak{F}_1$ we get $f = \sum_{i=1}^{k} L_i(f) g_i$. This shows that $\varphi(\mathfrak{N}(f)) = Sf$ is linear and interpolatory. Since $e(\varphi) = 0$, φ is an optimal error algorithm. ■

Without loss of generality we shall consider throughout this section information operators

$$\mathfrak{N} = [L_1, L_2, \ldots, L_n]^t \tag{8.3}$$

where $n = \operatorname{card}(\mathfrak{N}) < \dim \mathfrak{F}_1$. In Section 4 we showed that unless $\ker \mathfrak{N} \cap \ker T$ is contained in $\ker S$, $d(\mathfrak{N}, S, T) = +\infty$. This assumption holds if $\mathfrak{N}^* \subset \mathfrak{N}$ where

$$\mathfrak{N}^* = [L_1^*, L_2^*, \ldots, L_{n^*}^*]^t \text{ and } L_i^*(\xi_j^*) = \delta_{ij} \tag{8.4}$$

is defined by (6.1) with $n^* = \operatorname{index}(S, T) < +\infty$. Therefore we assume that $\mathfrak{N}$ defined by (8.3) satisfies

$$L_i = L_i^* \quad \text{for } i = 1, 2, \ldots, n^*, \; n \geq n^*. \tag{8.5}$$

Lemma 8.3

Suppose that there exist elements $g_1, g_2, \ldots, g_n$ from $\mathfrak{I}_1$ such that

(8.6) $L_i(g_j) = \delta_{ij}$ and $T(g_i) = 0$ for $i,j = 1,2,\ldots,n$.

Then $\varphi(\mathfrak{N}(f)) = \sum_{i=1}^{n} L_i(f) S g_i$ is a linear interpolatory optimal error algorithm and

$$e(\varphi) = r(\mathfrak{N},S,T) = \frac{1}{2} d(\mathfrak{N},S,T). \quad \blacksquare$$

Proof

Let $\tilde{f} = \sum_{j=1}^{n} L_j(f) g_j$. Then $L_i(\tilde{f}) = \sum_{j=1}^{n} L_j(f) L_i(g_j) = L_i(f)$ for $i = 1,2,\ldots,n$.

Thus $\mathfrak{N}(\tilde{f}) = \mathfrak{N}(f)$ and $\varphi(\mathfrak{N}(f)) = S\tilde{f}$ which proves that φ is a linear and interpolatory algorithm. Since $T(\tilde{f}) = 0$, Lemma 8.3 follows from Lemma 8.1. $\blacksquare$

Lemma 8.1 and 8.3 are not applicable unless there exists an element $\tilde{f}$ which shares the same information as f and belongs to ker T. We now show how to find a linear optimal (or close to optimal) error algorithm in general. Let

(8.7) $\mathfrak{I}_4 = T(\ker \mathfrak{N}) \oplus T(\ker \mathfrak{N})^{\perp}$

where $T(\ker \mathfrak{N})^{\perp} = \operatorname{lin}(\eta_1, \eta_2, \ldots, \eta_k)$. From Lemma 6.4 we know that $k = \dim T(\ker \mathfrak{N})^{\perp} \leq n - n^*$. Then for every $g \in \mathfrak{I}_4$ we have

(8.8) $g = g_0 + \sum_{i=1}^{k} R_i(g) \eta_i$

where $g_0 \in T(\ker \mathfrak{N})$ and $R_1, R_2, \ldots, R_k$ are linearly independent linear functionals such that $R_i(\eta_j) = \delta_{ij}$. Define

(8.9) $$c = \sup_{g \in \mathfrak{I}_4} \frac{\|g_0\|}{\|g\|}.$$

Note that c depends on ker $\mathfrak{N}$ and T but is independent of $\mathfrak{Z}_1$ and S. Furthermore $c \geq 1$ and if $T(\ker \mathfrak{N})$ is closed then c is finite. If $\mathfrak{Z}_4$ is a Hilbert space then we can assume that $T(\ker \mathfrak{N})^{\perp}$ is the orthogonal complement of $T(\ker \mathfrak{N})$, $(\eta_i, \eta_j) = \delta_{ij}$, $R_i(g) = (g, \eta_j)$ and $\|g\|^2 = \|g_0\|^2 + \sum_{i=1}^{k} |(g, \eta_i)|^2$ which implies $c = 1$.

Let $\mathfrak{Z}_1 = \ker \mathfrak{N} \oplus (\ker \mathfrak{N})^{\perp}$. Then $f = f_0 + \sum_{i=1}^{n} L_i(f)\xi_i$ where $f_0 \in \ker \mathfrak{N}$ and $L_i(\xi_j) = \delta_{ij}$ for $i,j = 1,2,\ldots,n$, $\xi_i = \xi_i^*$ for $i = 1,2,\ldots,n^*$, where ξ_i^* is defined by (8.4). Note that $Tf = Tf_0 + \sum_{i=n^*+1}^{n} L_i(f)\xi_i$ since $T\xi_i^* = 0$. Thus

$$k = \dim \operatorname{lin}(T\xi_{n^*+1}, \ldots, T\xi_n).$$

There exist linearly independent elements $\xi_{n^*+1}^*, \ldots, \xi_{n-k}^*$ such that $T\xi_i^* = 0$ and $\xi_i^* \in \operatorname{lin}(\xi_{n^*+1}, \ldots, \xi_n)$, $i = n^*+1, \ldots, n-k$.

Let $m = n-k$. Since $\mathfrak{Z}_4 = T(\mathfrak{Z}_1)$ there exist $\xi_{m+1}^*, \ldots, \xi_n^*$ such that $\eta_i = T\xi_{m+i}^*$ for $i = 1,2,\ldots,k$. Define

$$M = L_i(\xi_j^*)) \qquad i,j = 1,2,\ldots,n. \tag{8.10}$$

We show that M is nonsingular. Indeed, let

$$c_1 L_i(\xi_1^*) + \ldots + c_n L_i(\xi_n^*) = 0 \quad \text{for } i = 1,2,\ldots,n.$$

Then $\xi = c_1\xi_1^* + \ldots + c_n\xi_n^* \in \ker \mathfrak{N}$ and $T\xi \in T(\ker \mathfrak{N})$. Since $T\xi_i^* = 0$ for $i = 1,2,\ldots,m=n-k$, we get $T\xi = c_{m+1}T\xi_{m+1}^* + \ldots + c_nT\xi_n^* = c_{m+1}\eta_1 + \ldots + c_n\eta_k \in T(\ker \mathfrak{N})^{\perp}$. This implies $c_{m+1} = \ldots = c_n = 0$. Hence $\xi = c_1\xi_1^* + \ldots + c_m\xi_m^* \in (\ker \mathfrak{N})^{\perp}$ which yields $c_1 = \ldots = c_m = 0$. This proves that M is nonsingular. Define

$$(8.11)\quad [g_1,g_2,\ldots,g_n]^t = (M^t)^{-1}[\xi_1^*,\xi_2^*,\ldots,\xi_n^*]^t.$$

Note that $L_i(g_j) = \delta_{ij}$ for $i,j = 1,2,\ldots,m$. We are ready to prove the main result of this section.

Theorem 8.1

Let $n^* = \text{index}(S,T) \le n = \text{card}(\mathfrak{N}) < +\infty$. Let $c, g_1, g_2, \ldots, g_n$ be defined by (8.9) and (8.11) respectively. Then

$$(8.12)\quad \varphi(\mathfrak{N}(f)) = \sum_{i=1}^{n} L_i(f) S g_i$$

is linear, interpolatory and an optimal error algorithm within a factor of c, i.e.,

$$(8.13)\quad r(\mathfrak{N},S,T) \le e(\varphi) \le \frac{c}{2}\, d(\mathfrak{N},S,T) \le c\ r(\mathfrak{N},S,T).$$ ■

Proof

Since $Tf = Tf_0 + \sum_{i=1}^{n} L_i(f) T\xi_i$ we get

$$(8.14)\quad R_j(Tf) = \sum_{i=1}^{n} L_i(f)\ R_j(T\xi_i) \quad \text{for } j = 1,2,\ldots,k.$$

Set $f = \xi_i^*$ for $i = 1,2,\ldots,n$ in (8.14). Since $T\xi_i^* = 0$ for $i \le n-k$ and $R_j(T\xi_i^*) = R_j(\eta_{i-m}) = \delta_{i-m,j}$ for $i > m$, we get

$$(8.15)\quad [0,I] = M_1 M$$

where 0 is the $k \times (n-k)$ zero matrix, I is the $k \times k$ unit matrix and $M_1 = (R_i(T\xi_j))$ $i = 1,2,\ldots,k$, $j = 1,2,\ldots,n$.

Since M is nonsingular we have

$$M_1 = [0,I]M^{-1}. \tag{8.16}$$

From (8.11) and (8.16) we get

$$[Tg_1,\ldots,Tg_n]^t = (M^t)^{-1}[0,\ldots,0,\ \eta_1,\eta_2,\ldots,\eta_k]^t,$$

$$Tg_i = \sum_{j=1}^{k} R_j(T\xi_i)\eta_j. \tag{8.17}$$

We are ready to prove optimality of $\varphi(\mathfrak{N}(f)) = \sum_{i=1}^{n} L_i(f)Sg_i$. Let $\tilde{f} = \sum_{i=1}^{n} L_i(f)g_i$ for $f \in \mathfrak{F}_0$. Since $L_i(g_j) = \delta_{ij}$, then $\mathfrak{N}(\tilde{f}) = \mathfrak{N}(f)$. Thus $\varphi(\mathfrak{N}(f)) = S\tilde{f}$ is an interpolatory algorithm. Let $h = f-\tilde{f}$, $h \in \ker \mathfrak{N}$. From (8.17), (8.14) and (8.8) we get

$$Th = Tf - \sum_{i=1}^{n} L_i(f)Tg_i = Tf - \sum_{i=1}^{n} L_i(f)\sum_{j=1}^{k} R_j(T\xi_i)\eta_j =$$

$$= Tf - \sum_{j=1}^{k}\left(\sum_{i=1}^{n} L_i(f)R_j(T\xi_i)\right)\eta_j = Tf - \sum_{j=1}^{k} R_j(Tf)\eta_j =$$

$$= (Tf)_0 \in T(\ker \mathfrak{N}).$$

Then $\|Th\| = \|(Tf)_0\| \le c\|Tf\| \le c$, due to (8.9). Thus

$$\|\varphi(\mathfrak{N}(f)) - Sf\| = \|Sh\| \le c \sup_{\substack{h\in\ker \mathfrak{N} \\ \|Th\|\le 1}} \|Sh\| \le \frac{c}{2}\, d(\mathfrak{N},S,T) \le cr(\mathfrak{N},S,T).$$

Since f is an arbitrary element of $\mathfrak{F}_0$, $e(\varphi) \le \frac{c}{2}\, d(\mathfrak{N},S,T)$ and of course $e(\varphi) \ge r(\mathfrak{N},S,T)$. Hence (8.13) is proven which completes the proof. ■

Remark 8.1

Theorem 2.2 assures us of the existence of an interpolatory algorithm whose error is within at most a factor of two of the optimal error. Theorem 8.1 shows how to construct a linear interpolatory algorithm whose error is within at most a factor of c of the optimal error. ■

Note that if c = 1 then φ defined by (8.12) is optimal. This leads to

Corollary 8.1

If $\mathfrak{I}_4$ is a Hilbert space then the algorithm φ defined by (8.12) is a linear interpolatory optimal error algorithm and

$$e(\varphi) = r(\mathfrak{N},S,T) = \| ST^{-1} \|_{T(\ker \mathfrak{N})}. \quad \blacksquare$$

Corollary 8.1 and Lemma 5.1 yield

Corollary 8.2

If $\mathfrak{I}_4$ is a Hilbert space then

$$d(\mathfrak{N},S,T) = 2r(\mathfrak{N},S,T). \quad \blacksquare$$

The problem of linear optimal error algorithms was considered by many authors including Bakhvalov [71], Bojanov [74] and [76], Micchelli and Rivlin [77] and Smolyak [65]. They assumed that the solution operator S is a linear functional and $\mathfrak{I}_0$ is a convex and balanced set. See also Golomb and Weinberger [59] for some related material and discussion.

9. COMPLEXITY FOR THE LINEAR CASE

We specify our model of computation for linear problems (S,T) and linear information operators $\mathfrak{N}$ as follows. (For the general case, see Section 3.)

Model of Computation for the Linear Case

(i) Let P be a given collection of primitives. We assume that the addition of two elements of $\mathfrak{Z}_2$, f+g, and the multiplication of an element of $\mathfrak{Z}_2$ by a scalar, cf, are primitive operations which belong to P. We also assume that every linear functional L, $L: \mathfrak{Z}_1 \to \mathbb{C}$, is a primitive operation which belongs to P. This implies that any linear information operator $\mathfrak{N} = [L_1, L_2, \ldots, L_n]^t$ of finite cardinality is permissible where $L_1, L_2, \ldots, L_n$ are arbitrary linear functionals.

(ii) To normalize the measure of the complexity we assume that the cost of the addition of two elements of $\mathfrak{Z}_2$ and the multiplication of an element of $\mathfrak{Z}_2$ by a scalar is taken as unity. Note that for a finite dimensional space $\mathfrak{Z}_2$, $m = \dim \mathfrak{Z}_2$, unit cost means the cost of m scalar additions or multiplications.

Let comp(L) denote the complexity of evaluating a linear functional L. Let $\mathfrak{N} = [L_1, L_2, \ldots, L_n]^t$ be a linear information operator with linearly independently linear functionals $L_1, L_2, \ldots, L_n$, $\text{card}(\mathfrak{N}) = n$. We assume that $\mathfrak{N}(f)$ is computed by the independent evaluation of $L_1(f), L_2(f), \ldots, L_n(f)$ and the information complexity, see (3.4), of $\mathfrak{N}$ is given by

$$(9.1)\quad \mathrm{comp}(\mathfrak{N}) = \sum_{i=1}^{n} \mathrm{comp}(L_i).$$

If $\mathrm{comp}(L_i) \equiv c_1$ then $\mathrm{comp}(\mathfrak{N}) = nc_1$ which shows how the information complexity depends on the cardinality of $\mathfrak{N}$.

(iii) Let φ be a permissible algorithm which uses $\mathfrak{N}(f)$ and finds an ε-approximation to $\alpha = S(f)$. Let $d(\varphi)$ be the combinatory complexity of φ. For all problems of practical interest, φ has to use every $L_i(f)$, $i = 1,2,\ldots,n$, at least once and $d(\varphi) \geq n-1$. We rule out special problems and information operators, assuming that $d(\varphi) \geq n-1$ for every algorithm under consideration. ■

Example 9.1

Let $P = \{$arithmetic operations, the evaluation of linear functionals, the evaluation of a function, the evaluation of a derivative$\}$. Let $\mathfrak{N}(f) = \{f(x), f'(x)\}$ where f is a function in a space of dimension m, $m \leq \infty$. Assume $\varphi(\mathfrak{N}(f)) = x - (f'(x))^{-1} f(x)$.

Case 1. $m < \infty$. Let $\mathrm{comp}(L) = c_1$, for every linear functional L. Then $\mathrm{comp}(f(x)) = mc_1$, $\mathrm{comp}(f'(x)) = m^2 c_1$. We compute $\varphi(\mathfrak{N}(f))$ by solving the appropriate linear system by Gaussian elimination. Thus, by (ii), $d(\varphi) = O(m^2)$ times unit cost. We conclude

$$\mathrm{comp}(\varphi) = mc_1 + m^2 c_1 + d(\varphi).$$

Case 2. $m = \infty$. We add the solution of a linear system to our set of primitives; let the complexity of this primitive be c_4. Let $\mathrm{comp}(f(x)) = c_2$, $\mathrm{comp}(f'(x)) = c_3$. Then $\mathrm{comp}(\varphi) = c_2 + c_3 + c_4$. ■

This model of computation is an idealized one since we assume that every linear functional is a primitive operation. However, even in this idealized model we shall prove that the complexity of a linear problem (S,T) as a function of ϵ can be essentially any decreasing function of ϵ.

Let φ be a linear algorithm, i.e. $\varphi(\mathfrak{N}(f)) = \sum_{i=1}^{n} L_i(f)g_i$ for certain elements $g_1, g_2, \ldots, g_n$ from $\mathfrak{Z}_2$. The element g_i depends on $\mathfrak{N}$, S and T but is independent of f. Therefore the elements $g_1, g_2, \ldots, g_n$ can be precomputed. The computation of $\varphi(\mathfrak{N}(f))$, given $\mathfrak{N}(f)$, requires at most n multiplications and n-1 additions which are primitive operations with unit cost. Thus, any linear algorithm φ is permissible and its combinatory complexity is at most 2n-1. Due to (iii), every linear algorithm is within a factor of 2 of minimal combinatory complexity. Therefore to make combinatory complexity small it is desirable to make the cardinality of $\mathfrak{N}$ as small as possible under the constraint that $r(\mathfrak{N},S,T) < \epsilon$.

Fix $\mathfrak{N}$, S and T. Recall that $\Phi(\epsilon)$ is the class of <u>all</u> permissible algorithms φ which use the information operator $\mathfrak{N}$ and whose error is not larger than ϵ, $e(\varphi) \leq \epsilon$.

<u>Lemma 9.1</u>

Let $r(\mathfrak{N},S,T) < \epsilon$. Assume there exists a linear algorithm φ in $\Phi(\epsilon)$. Then,

$$(9.2)\quad \left(1 + \frac{n-1}{\mathrm{comp}(\mathfrak{N})}\right) \Big/ \left(1 + \frac{2n-1}{\mathrm{comp}(\mathfrak{N})}\right) \leq \frac{\mathrm{comp}(\mathfrak{N},S,T,\epsilon)}{\mathrm{comp}(\varphi)} \leq 1. \qquad \blacksquare$$

Proof

Since a linear algorithm is permissible, $\Phi(\epsilon)$ is not empty. The complexity of a linear algorithm φ, $\mathrm{comp}(\varphi)$, is not larger than $\mathrm{comp}(\mathfrak{N}) + 2n-1$. Since $\mathrm{comp}(\mathfrak{N},S,T,\epsilon) \geq \mathrm{comp}(\mathfrak{N}) + n-1$, (9.2) is proven. ■

Note that if $\mathrm{comp}(\mathfrak{N}) \geq n$ then (9.2) yields

$$(9.3)\quad \frac{1}{2} \leq \frac{2n-1}{3n-1} \leq \frac{\mathrm{comp}(\mathfrak{N},S,T,\epsilon)}{\mathrm{comp}(\varphi)} \leq 1.$$

In many cases $\mathrm{comp}(\mathfrak{N}) \gg n$ which yields

$$(9.4)\quad \mathrm{comp}(\mathfrak{N},S,T,\epsilon) \cong \mathrm{comp}(\varphi) \cong \mathrm{comp}(\mathfrak{N})$$

for every linear algorithm φ.

Fix S,T, and let ϵ tend to zero. We seek the minimal cardinality of an information operator $\mathfrak{N}$ such that $r(\mathfrak{N},S,T) < \epsilon$. Let Ψ be a class of permissible information operators such that $\inf_{\mathfrak{N}\in\Psi} r(\mathfrak{N},S) = 0$. See (3.8). Define

$$(9.5)\quad m(\Psi,S,T,\epsilon) = \min\{\mathrm{card}(\mathfrak{N}):\ \mathfrak{N} \in \Psi,\ r(\mathfrak{N},S,T) < \epsilon\}.$$

We shall say $m(\Psi,S,T,\epsilon)$ is the <u>ϵ-cardinality number for the problem (S,T) in the class Ψ</u>. Note that $m(\Psi,S,T,\epsilon)$ is a nonincreasing function of ϵ.

Let Ψ_U be the class of <u>all</u> linear information operators $\mathfrak{N}$ such that $\mathrm{card}(\mathfrak{N}) < +\infty$. ($\mathfrak{N}$ is permissible due to (i)!) Note that $\bigcup_{n=n^*}^{\infty} \Psi_n \subset \Psi_U$ where Ψ_n, see Section 6, is the class of all information operators $\mathfrak{N}$ such that $\mathfrak{N}^* \subset \mathfrak{N}$, $\mathrm{card}(\mathfrak{N}) \leq n$. Furthermore $\bigcup_{n=0}^{\infty} \Psi_n = \Psi_U$ if $\mathrm{index}(S,T) = 0$.

The class Ψ_U contains all information operators of practical interest since every information operator which is to be computed has to have finite cardinality. Therefore the ϵ-complexity $\mathrm{comp}(\Psi_U,S,T,\epsilon)$ is the <u>inherent complexity of the problem (S,T)</u>.

We show that $m(\Psi_U, S, T, \epsilon)$ can be essentially <u>any</u> decreasing function of ϵ. More precisely assume that ϵ belongs to the interval $(0, \epsilon_0]$. Let

(9.6) $g: (0, \epsilon_0] \to R^+$

be a decreasing function such that $g(\epsilon_0) \geq 1$ and $\lim_{\epsilon \to 0^+} g(\epsilon) = +\infty$.

<u>Theorem 9.1</u>

For every function g defined by (9.6) there exists a linear problem (S,T) such that

(9.7) $g(\epsilon) - 1 < m(\Psi_U, S, T, \epsilon) \leq g(\epsilon), \quad \forall \epsilon \in (0, \epsilon_0]$.

Furthermore there exists a sequence $\{\epsilon_i\}$ such that $\epsilon_i \in (0, \epsilon_0]$, $\lim_{i \to \infty} \epsilon_i = 0$ and

(9.8) $m(\Psi_U, S, T, \epsilon_i) = g(\epsilon_i)$. ■

<u>Proof</u>

Let $g^{-1}: [g(\epsilon_0), +\infty) \to R^+$ be the inverse function of g. Define $\beta_i = \epsilon_0 + 1$ for $i < g(\epsilon_0)$ and $\beta_i = g^{-1}(i)$ for $i \geq g(\epsilon_0)$. Note that $\lim_{i \to \infty} \beta_i = 0$.

Let $\mathcal{Y}_4 = \mathcal{Y}_2 = \text{lin}(\xi_1, \xi_2, \ldots)$ be an infinite dimensional Hilbert space where $(\xi_i, \xi_j) = \delta_{ij}$. Define $T = I$ and

(9.9) $Sf = \sum_{i=1}^{\infty} \beta_i (f, \xi_i) \xi_i$.

Thus S is a self-adjoint and compact operator. Furthermore $Sf_i = \beta_i \xi_i$ for $i = 1, 2, \ldots$.

Note that $n^* = 0$ where $n^* = \text{index}(S, I)$. From (7.2) we get $K_1 = S^2$ and the eigenvalues of K_1 satisfy $\lambda_i = \beta_i^2$ for $i = 1, 2, \ldots$.

Let $m = m(\Psi_U, S, I, \varepsilon)$. This means there exists an information operator $\mathcal{N}_0$ such that $\text{card}(\mathcal{N}_0) = m$ and $r(\mathcal{N}_0, S, T) < \varepsilon$. Moreover for every $\mathcal{N}$ such that $\text{card}(\mathcal{N}) < m$, $r(\mathcal{N}, S, T) \geq \varepsilon$. Due to Theorem 7.2 and Corollary 8.1 we know that

$$\beta_{m+1} \leq r(\mathcal{N}_0, S, T) < \varepsilon.$$

Thus $m+1 \geq g(\varepsilon_0) + 1$ and $\beta_{m+1} = g^{-1}(m+1) < \varepsilon$ which yields $m > g(\varepsilon) - 1$. Furthermore for $\mathcal{N} = \mathcal{N}_{m-1}$ defined by (7.5) we get $\text{card}(\mathcal{N}_{m-1}) < m$ and $r(\mathcal{N}_{m-1}, S, I) = \beta_m \geq \varepsilon$. This yields $m \leq g(\varepsilon)$ and proves (9.7).

Let $\varepsilon_i = \beta_{i+1}$ for $i \geq g(\varepsilon_0)$. Then $\varepsilon_i \in (0, \varepsilon_0]$ and $\lim_i \varepsilon_i = 0$. Since $r(\mathcal{N}_i, S, I) = \beta_{i+1} = \varepsilon_i$ we get $m(\Psi_U, S, I, \varepsilon) = i+1 = g(\varepsilon_i)$. This proves (9.8) and completes the proof. ■

Theorem 9.1 states that $m(\Psi_U, S, T, \varepsilon)$ can be an essentially arbitrary function of ε. Recall that the ε-complexity $\text{comp}(\Psi_U, S, T, \varepsilon)$ of the problem (S,T) in the class Ψ_U is defined by (3.9). From Theorem 9.1 we can conclude that $\text{comp}(\Psi_U, S, T, \varepsilon)$ can depend arbitrarily on ε. To show this, assume for the sake of simplicity that the complexity of evaluating any linear functional L_i is fixed, $\text{comp}(L_i) = c_1$.

Theorem 9.2

For every function g defined by (9.6) there exists a linear problem S,T such that

$$(9.10)\quad g(\varepsilon)(c_1+1) - c_1 - 2 < \text{comp}(\Psi_U, S, T, \varepsilon) \leq g(\varepsilon)(c_1 + 2) - 1, \quad \forall \varepsilon \in (0, \varepsilon_0). \blacksquare$$

Proof

Consider the problem S,T defined in the proof of Theorem 9.1. Thus, the ε-cardinality number $m = m(\Psi_U, S, T, \varepsilon)$ satisfies (9.7) and the information complexity $\mathcal{N}$ such that $\text{card}(\mathcal{N}) = m$, $r(\mathcal{N}, S, T) < \varepsilon$, satisfies

(9.11) $(g(\varepsilon) - 1)c_1 < \mathrm{comp}(\mathfrak{N}) = m(\Psi_U, S, T, \varepsilon)c_1 \leq g(\varepsilon)c_1$.

Since the problem (S,T) is defined in a Hilbert space, Corollary 8.1 assures the existence of a linear optimal error algorithm φ, $e(\varphi) = r(\mathfrak{N}, S, T) < \varepsilon$. Thus

(9.12) $m-1 + \mathrm{comp}(\mathfrak{N}) \leq \mathrm{comp}(\Psi_U, S, T, \varepsilon) \leq \mathrm{comp}\ \varphi \leq 2m-1 + \mathrm{comp}(\mathfrak{N})$.

From (9.11) and (9.12) we get (9.10). ■

Theorem 9.2 states that $\mathrm{comp}(\Psi_U, S, T, \varepsilon)$ is roughly the same function of ε as the ε-cardinality number $m(\Psi_U, S, T, \varepsilon)$. Note that the function g can tend to infinity arbitrarily fast as ε tends to zero. This proves

Corollary 9.1

(i) There exist linear problems with arbitrarily hard complexity.

(ii) There are no "gaps" in the complexity functions. ■

This may be contrasted with the theory of recursively computable functions where complexity gaps are known to occur (Borodin [72]).

Remark 9.1

We assumed that $\mathfrak{N}$ consists of linear functionals which are computed independently and therefore $\mathrm{comp}(\mathfrak{N}) = nc_1$. For some information operators, $\mathfrak{N}(f)$ can be computed faster than nc_1. For instance assume that $L_i(f) = f(x_i)$ for distinct points x_i, $i = 1,\ldots,n$, where f is a polynomial of degree n-1. Then the complexity of L_i is $O(n)$ but $\mathfrak{N}(f)$ can be computed in $O(n \log^2 n)$. In fact, Theorem 9.2 remains valid under the relaxed assumption that $\mathrm{comp}(\mathfrak{N}(f)) = \omega(n)$ where ω is an increasing function of n with $\lim_{n\to\infty} \omega(n) = +\infty$. ■

CHAPTER III

APPLICATIONS

In this section we apply the general theory to a variety of problems. We consider rather simple problems since we wish to show the importance and usefulness of the ideas and avoid overwhelming the reader with technical details.

10. APPROXIMATION IN A HILBERT SPACE

Let $H = \text{lin}(\xi_1, \xi_2, \ldots)$ be an infinite dimensional Hilbert space over the real field R where $(\xi_i, \xi_j) = \delta_{ij}$. Thus $f \in H$ iff $f = \sum_{i=1}^{\infty} (f, \xi_i)\xi_i$ and $\sum_{i=1}^{\infty} (f, \xi_i)^2 < \infty$. Let $\{\beta_i\}$ be a nonzero sequence of real numbers such that $|\beta_i| \leq |\beta_{i+1}|$ for all i. Let

$$(10.1) \quad \mathfrak{I}_1 = \mathfrak{I}_2 = \{f: f \in H \text{ and } \sum_{i=1}^{\infty} \beta_i^2 (f, \xi_i)^2 < +\infty\}.$$

We consider an approximation problem defined by

$$(10.2) \quad Sf = f, \quad f \in \mathfrak{I}_0$$

where $\mathfrak{I}_0 = \{f: f \in \mathfrak{I}_1 \text{ and } \|Tf\| \leq 1\}$ for

$$(10.3) \quad Tf = \sum_{i=1}^{\infty} \beta_i (f, \xi_i)\xi_i, \quad \mathfrak{I}_4 = T(\mathfrak{I}_1) \subset H.$$

We first find the index $n^* = \text{index}(I,T)$. Since $\ker S = \{0\}$, (5.3) yields $A(T,S) = \ker T$ and $n^* = \dim(\ker T)$. Let i_0 be the largest index such that $\beta_i = 0$ for $i = 1,2,\ldots,i_0$ $(\beta_{i_0+1} \neq 0)$. Then $n^* = i_0 < +\infty$. Note that $\mathfrak{N}^*(f) = [(f,\xi_1),(f,\xi_2),\ldots,(f,\xi_{n^*})]^t$ satisfies Lemma 6.1, i.e.,

$\ker \Lambda^* \cap T = \{0\} = \ker S$. From (6.3) we easily find the inverse of T^{-1},

$$T^{-1}f = \sum_{i=n^*+1}^{\infty} \frac{1}{\beta_i} (f,\xi_i)\xi_i$$

and $K = ST^{-1} = T^{-1}$. Note that K is self adjoint and $K\xi_i = \frac{1}{\beta_i}\xi_i$. Thus K is compact iff $\lim_i \beta_i = +\infty$. From Theorem 7.1 it follows that the problem (I,T) is convergent iff $\lim_i \beta_i = +\infty$.

We want to find the n-th optimal information Λ_n for $n \geq n^*$. Since $K(\mathcal{Y}_4) \subset \mathcal{Y}_4$, we get from (7.5) that

$$(10.4) \quad \Lambda_n(f) = [(f,\xi_1),(f,\xi_2),\ldots,(f,\xi_n)]^t$$

is the n-th optimal information and due to Theorem 7.2 and Corollary 8.1 we have

$$(10.5) \quad r(\Lambda_n,S,T) = \frac{1}{|\beta_{n+1}|} .$$

The linear optimal error algorithm φ defined in Theorem 8.1 is given by

$$(10.6) \quad \varphi(\Lambda_n(f)) = \sum_{i=1}^{n} (f,\xi_i)\xi_i, \quad e(\varphi) = \frac{1}{|\beta_{n+1}|} .$$

Note that φ is the initial section of $f = \sum_{i=1}^{\infty} (f,\xi_i)\xi_i$.

We analyze the complexity of this problem. Recall that Ψ_U is the class of <u>all</u> linear information operators with finite cardinality. It is easy to see that the ε-cardinality number $m = m(\Psi_U,I,T,\varepsilon)$, see (9.5), is equal to the smallest number n such that $\frac{1}{|\beta_{n+1}|} < \varepsilon$. Since there exists a linear optimal error algorithm for any ε we get bounds on the ε-complexity of the problem (S,T) in the class Ψ_U,

(10.7) $(c_1+1)m(\Psi_U, I, T, \epsilon) - 1 \leq \text{comp}(\Psi_U, I, T, \epsilon) \leq (c_1+2)m(\Psi_U, I, T, \epsilon) - 1$

where c_1 is the complexity of evaluating a linear functional of the form (f, ξ_i). Note that m depends only on how fast β_i goes to infinity and due to Theorem 9.1, $m = m(\Psi_U, I, T, \epsilon)$ can be essentially any decreasing function of ϵ.

11. UNIFORM APPROXIMATION

Let k be an nonnegative integer. Let $\mathfrak{F}_1 = C[0,1]$ for $k = 0$ and $\mathfrak{F}_1 = \{f: f \in C^{k-1}[0,1],\ f^{(k)} \in L_\infty[0,1]\}$ with $\|f\| = \max_{0\le t\le 1} |f(t)|$. Let $\mathfrak{F}_2 = \mathfrak{F}_1$ and $\mathfrak{F}_4 = L_\infty[0,1]$. Define

$$(11.1)\quad Sf = f, \quad Tf = \frac{1}{k!} f^{(k)}.$$

Since $\ker T = \{f: f^{(k)} \equiv 0\}$, we obtain $n^* = \dim(\ker T) = k$. Consider an information operator

$$(11.2)\quad \mathfrak{N}(f) = [f(t_1), f(t_2), \ldots, f(t_n)]^t$$

for distinct $t_i \in [0,1], i = 1,2,\ldots n$ with $n \ge k$. Note that $\ker \mathfrak{N} \cap \ker T = \{0\} = \ker S$.

For $k = 0$, Example 6.1 yields $r(\mathfrak{N}, I, I) = 1$ for any information operator with finite cardinality. To assure convergence of the problem (I,T) we assume $k \ge 1$. We consider two cases.

Case 1, $n = k$.

Let $\omega(t) = \prod_{i=1}^{n} (t-t_i)$ and $g_i(t) = \omega(t)/\{(t-t_i)\omega'(t_i)\}$ for $i = 1,2,\ldots,n$. Observe that $g_i(t_j) = \delta_{ij}$ and $Tg_i \equiv 0$. From Lemma 8.3 we know that

$$(11.3)\quad \varphi(\mathfrak{N}(f)) = \sum_{i=1}^{n} f(t_i) g_i(t)$$

is a linear interpolatory optimal error algorithm. Furthermore $Sf(t) - \varphi(\mathfrak{N}(f)) = g(t)\omega(t)$ where $g(t) = f(t_1, t_2, \ldots, t_n; t)$ is the n-th divided difference of f and $\|g\| \le 1$. This yields

(11.4) $e(\varphi) = r(\mathfrak{N},I,T) = \|\omega\|$.

We seek optimal points $t_1^*,\ldots,t_n^*$ which minimize $e(\varphi) = e(\varphi,t_1,t_2,\ldots,t_n)$. It is well-known that $t_1^*,\ldots,t_n^*$ are uniquely defined by the zeros of the Chebyshev polynomial $T_n(\frac{1}{2}(t+1))$, i.e., $t_i^* = 2\cos(\frac{\pi}{2n} + \frac{i-1}{n}\pi) - 1$ for $i = 1,2,\ldots,n$ and

(11.5) $e(\varphi,t_1^*,\ldots,t_n^*) = \frac{2}{4^n}$.

Case 2, $n > k$.

Micchelli, Rivlin and Winograd [76] show there exists a linear optimal error algorithm φ such that $e(\varphi) = r(\mathfrak{N},I,T)$ and

(11.6) $r(\mathfrak{N},I,T) = \|q\|$

where q is a perfect spline of degree k having at most $n-k$ knots and $q(t_i) = 0$ for $i = 1,2,\ldots,n$, $\|\frac{q^{(k)}}{k!}\| = 1$. Furthermore if $t_1 = 0$, $t_n = 1$ and $\Delta_n = \max_{1\le i\le n-1} (t_{i+1}-t_i)$ they prove

(11.7) $\frac{\Delta_n^k}{4^k} \le r(\mathfrak{N},I,T) \le \frac{k!}{4^k}\Delta_n^k$.

Furthermore the points $t_1^*,t_2^*,\ldots,t_n^*$ which minimize $r(\mathfrak{N},I,T) = r(\mathfrak{N},I,T;\ t_1,t_2,\ldots,t_n)$ are the zeros of the Chebyshev perfect spline of degree k with $n-k$ knots and $r(\mathfrak{N},I,T;\ t_1^*,\ldots,t_n^*)$ is the n-width of $\mathfrak{I}_0$ in the sense of Gelfand (Tichomirov [69]) and

(11.8) $r(\mathfrak{N},I,T,t_1^*,t_2^*,\ldots,t_n^*) = c_{k,n}h^k$, $\forall\ n > k$,

where $0 < \underline{c}_k \le c_{k,n} \le \bar{c}_k < +\infty$ for some constants $\underline{c}_k$ and $\bar{c}_k$, $h = \frac{1}{n}$. ■

We analyze the complexity of uniform approximation. Let Ψ be the class of all information operators $\mathfrak{N}$ such that $\mathfrak{N}(f) = [f(t_1), f(t_2), \ldots, f(t_n)]^t$ for any distinct t_i and any n. If $\frac{2}{4^k} < \varepsilon$ then choose $n = k$ and (11.5) assures that the ε-complexity of the problem (I,T) in the class Ψ satisfies

$$(11.9) \quad (c_1+1)k-1 \le \mathrm{comp}(\Psi,I,T,\varepsilon) \le (c_1+2)\cdot k-1$$

where c_1 is the complexity of one function evaluation. If $\frac{2}{4^k} \ge \varepsilon$, let $n = n(\varepsilon)$ be the smallest integer such that

$$(11.10) \quad n > \frac{\sqrt[k]{c_{k,n}}}{\sqrt[k]{\varepsilon}} .$$

Due to (11.8) we get $\sqrt[k]{c_k}/\sqrt[k]{\varepsilon} < n(\varepsilon) \le \sqrt[k]{c_k}/\sqrt[k]{\varepsilon} + 1$. We have

$$(11.11) \quad (c_1+1)n(\varepsilon)-1 \le \mathrm{comp}(\Psi,I,T,\varepsilon) \le (c_1+2)n(\varepsilon) - 1.$$

Note that (11.11) is tight for $c_1 \gg 1$. The conditions (11.10) and (11.11) state that

$$\mathrm{comp}(\Psi,I,T,\varepsilon) = 0\left(\frac{1}{\sqrt[k]{\varepsilon}}\right).$$

This shows that asymptotically in ε, complexity decreases as the regularity of the class of problem elements increases.

12. EVALUATION OF INTEGRALS

Let k be a nonnegative integer. For $k = 0$ let $\mathfrak{F}_1 = C[0,1]$ and for $k \geq 1$ let $\mathfrak{F}_1 = \{f: f^{(k)}$ is piecewise continuous on $[-1,1]\}$. Define $\mathfrak{F}_2 = R$, $\mathfrak{F}_4 = L_2[-1,1]$ and

$$(12.1) \quad Sf = \int_{-1}^{1} f(t)dt \text{ and } Tf = \frac{1}{k!} f^{(k)}.$$

Consider the information operator $\mathfrak{N}$ given by

$$(12.2) \quad \mathfrak{N}(f) = [f(t_1),\ldots,f^{(j-1)}(t_1),\ldots,f(t_m),\ldots,f^{(j-1)}(t_m)]^t$$

for $-1 \leq t_1 < t_2 < \ldots < t_m \leq 1$ where $j = 1$ for $k = 0$ and $j-1 < k$ for $k \geq 1$. Note that $\text{card}(\mathfrak{N}) = jm$. Since $n^* = \text{diam } A(T,S) = k - \left\lceil \frac{k-1}{2} \right\rceil$ we have to assume $jm \geq k - \left\lceil \frac{k-1}{2} \right\rceil$.

For $k = 0$ it is easy to verify that $r(\mathfrak{N},S,I) = \sqrt{2}$ no matter how many function evaluations are known. To assure convergence of the problem (S,T) we have to assume $k \geq 1$. For $k = j$ this problem was considered by Bojanov [76] who showed the existence of a linear optimal error algorithm φ. His analysis is based on Smolyak's lemma (Smolyak [65], Bakhvalov [71]). Note that Theorem 8.1 and Corollary 8.1 assure the existence of a linear optimal error algorithm defined by (8.12) for any j and k.

For the sake of simplicity we assume here that $k = j$. Then the error $e(\varphi) = r(\mathfrak{N},S,T)$ is given by Bojanov,

$$(12.3) \quad r(\mathfrak{N},S,T) = \left\{ \frac{(t_1-1)^{2k+1}}{2k+1} + \frac{(1-t_m)^{2k+1}}{2k+1} + E_k^2 \sum_{i=1}^{m-1} \left(\frac{t_{i+1}-t_i}{2} \right)^{2k+1} \right\}^{\frac{1}{2}}$$

where E_k is the minimal error of approximation of the function t^k by

polynomials of lower degree in the space L_2. From Akhieser [56, p. 19] we get

$$(12.4)\quad E_k = \frac{(k!)^2}{\sqrt{2k+1}(2k!)} = \sqrt{\frac{\pi}{2}}\ \frac{1}{4^k}\ (1 + o(k)).$$

How should the points $t_1, t_2, \ldots, t_m$ be chosen so as to minimize the radius $r(\mathcal{A}, S, T)$? Bojanov shows

$$(12.5)\quad \min_{t_i} r(\mathcal{A}, S, T, t_1, \ldots, t_m) = h_m^k E_k$$

where $h_m = \left(m-1 + \sqrt[2k]{\frac{2k+1}{2}} \cdot \sqrt[k]{E_k}\right)^{-1} = \frac{1}{m}(1 + o(m))$ and the optimal points are equal to

$$(12.6)\quad t_i^* = -1 + (E_k^2(2k+1)/2)^{1/(2k)} \cdot h_m + 2(i-1)h_m.$$

Note that $t_1^*, \ldots, t_m^*$ are equally spaced.

Observe that commonly used Gauss quadrature is not an optimal error algorithm since it is based on the zeros of orthogonal polynomials which are not equally spaced.

We analyze the complexity of evaluation of integrals. Let Ψ be the class of <u>all</u> information operators of the form (12.2) for any distinct t_i, any m, and with k = j. Let $m = m(\varepsilon)$ be the smallest integer such that

$$(12.7)\quad h_m^k E_k < \varepsilon.$$

For small ε, $m = \frac{1}{4\varepsilon^{1/k}}(1 + o(1))$. From (12.7) we get bounds on the ε-complexity $\mathrm{comp}(\Psi, S, T, \varepsilon)$,

$$(12.8)\quad m(\varepsilon) \sum_{i=1}^{k} c(f^{(i-1)}) + km(\varepsilon) - 1 \le \mathrm{comp}(\Psi, S, T, \varepsilon) \le m(\varepsilon) \sum_{i=1}^{k} c(f^{(i-1)}) + 2km(\varepsilon) - 1$$

where $c(f^{(i-1)})$ is the complexity of one evaluation of $f^{(i-1)}$. For small ϵ and $c(f^{(i-1)}) \equiv c_1 \gg 1$, (12.8) becomes

$$(12.9) \quad \mathrm{comp}(\Psi,S,T,\epsilon) \cong \frac{k}{4\epsilon^{1/k}}\, c_1.$$

This shows that asymptotically in ϵ, the ϵ-complexity $\mathrm{comp}(\Psi,S,T,\epsilon)$ is a decreasing function of the regularity parameter k.

We conclude this section by the following remark. It is well known that it is impossible to find an ϵ-approximation to the integral of f knowing only the values of f and/or its derivatives at arbitrary but finite number of points. The argument is that adding, for instance, the polynomial $\omega(t) = c \prod_{i=1}^{m} (t-t_i)^{2j}$ to a function f, one does not change the information on f, $\mathcal{N}(f) = \mathcal{N}(f+\omega)$, but the value of integral $\int_{-1}^{1}(f(t) + \omega(t))dt$ can be arbitrarily different from the integral of f. In our setting we rule out this argument since we consider functions whose kth normalized derivative is bounded by one. We show that provided $k \geq 1$, then the integration problem can always be solved to within any ϵ and that at worst the complexity goes as $1/\epsilon$.

Compare with the result of Bojanov [74] who shows that for analytic f and $T = I$, $\mathrm{comp}(\Psi,S,T,\epsilon) = O(c_1(\log \frac{1}{\epsilon})^2)$.

13. EVALUATION OF A FUNCTION AND BOUNDARY VALUE PROBLEMS

Let $\mathcal{F}_1 = \mathcal{F}_2 = \{f: f \text{ is continuous on } [0,1] \text{ and } f' \in L_\infty(0,1)\}$ with the uniform norm $\|f\| = \max_{1 \le x \le 1} |f(x)|$. Let $\mathcal{F}_4 = L_\infty(0,1)$. Define

$$(13.1) \quad (Sf)(x) = \int_0^1 G(x,t) f(t)\,dt, \quad Tf = f'$$

where $G \ge 0$, $G \not\equiv 0$ is a continuous function on $[0,1] \times [0,1]$. The information operator $\mathcal{N}$ is given by

$$(13.2) \quad \mathcal{N}(f) = [f(t_0), f(t_1), \ldots, f(t_{n+1})]^t$$

where $t_i = ih$, $h = 1/(n+1)$ and $\text{card}(\mathcal{N}) = n+2$.

Thus we want to approximate the function (13.1) knowing the value of the integrand f at some points and the bound $\|f'\| \le 1$. Define

$$(13.3) \quad \tilde{f}(t) = \begin{cases} f(t_0) & 0 \le t < h/2, \\ f(t_i) & t_i - h/2 \le t < t_i + h/2, \ i = 1,2,\ldots,n, \\ f(t_{n+1}) & t_{n+1} - h/2 \le t \le 1. \end{cases}$$

Thus $\tilde{f}$ is a piecewise constant function. Note that $\tilde{f}(t_i) = f(t_i)$ and $Tf = 0$ almost anywhere. Since $\tilde{f}$ does not belong to $\mathcal{F}_0$ we cannot use Lemma 8.3. However we can still show that the linear algorithm

$$(13.4) \quad \varphi(\mathcal{N}(f)) = S\tilde{f} = f(t_0) \int_0^{h/2} G(x,t)\,dt + f(t_{n+1}) \int_{1-h/2}^1 G(x,t) + \sum_{i=1}^n f(t_i) \int_{t_i - h/2}^{t_i + h/2} G(x,t)\,dt$$

is an optimal error algorithm. Indeed

$$Sf - \varphi(\mathfrak{N}(f)) = \sum_{i=0}^{n} \left[\int_{t_i}^{t_i+h/2} G(x,t)(f(t)-f(t_i))dt + \int_{t_i+h/2}^{t_{i+1}} G(x,t)(f(t)-f(t_{i+1}))dt \right].$$

Since $f(t) - f(t_i) = f(t,t_i)(t-t_i)$, with the first divided difference $|f(t,t_i)| \le 1$, we get

$$(13.6) \quad \| Sf-\varphi(\mathfrak{N}(f)) \| \le \| \int_0^1 G(x,t)q(t)dt \| = \| Sq \|$$

where

$$(13.7) \quad q(t) = \begin{cases} t-t_i & t_i \le t \le t_i+h/2 \\ t_{i+1}-t & t_i+h/2 \le t \le t_{i+1}. \end{cases}$$

Note that $q(t_i) = 0$ for $i = 0,1,\ldots,n+1$ and $Tq = \pm 1$ almost everywhere. Thus $q \in \ker \mathfrak{N}$ and $\| Tq \| \le 1$. From Lemma 5.1 we get

$$\| Sf-\varphi(\mathfrak{N}(f)) \| \le \frac{1}{2} d(\mathfrak{N},S,T) \le r(\mathfrak{N},S,T)$$

which proves optimality of φ. It is easy to verify that

$$(13.8) \quad e(\varphi) = r(\mathfrak{N},S,T) = \| Sq \| = c_n h$$

where $c_n = c_n(G)$ and there exist two constants $\underline{c} = \underline{c}(G) > 0$ and $\bar{c} = \bar{c}(G) < +\infty$ such that $\underline{c} \le c_n \le \bar{c}$ for all n.

We analyze the complexity of evaluation of (13.1). Let Ψ be the class of <u>all</u> information operators of the form (13.2) for some n. Let $n = n(\epsilon)$ be the smallest integer such that

$$(13.9) \quad \frac{c_n}{n+1} < \epsilon.$$

Due to (13.8) we get $\underline{c}(G)/\epsilon < n(\epsilon) + 1 \le \bar{c}(G)/\epsilon + 1$. Then the ϵ-complexity $\mathrm{comp}(\Psi,S,T,\epsilon)$ satisfies

(13.10) $(c_1+1)(n(\varepsilon)+2) - 1 \le \mathrm{comp}(\Psi,S,T,\varepsilon) \le (c_1+2)(n(\varepsilon)+2) - 1$

where c_1 is the complexity of one function evaluation. For small ε and $c_1 \gg 1$, (13.10) becomes

(13.11) $\mathrm{comp}(\Psi,S,T,\varepsilon) \cong \frac{c_1}{\varepsilon}$.

It is possible to generalize this result and to show that if $\mathfrak{I}_1 = \{f: f \in C^{k-1}(0,1)$ and $f^{(k)} \in L_\infty(0,1)\}$ and $Tf = \frac{1}{k!} f^{(k)}$ there exists a linear optimal error algorithm φ which uses the information (13.2) with $n \ge k$ such that

(13.12) $e(\varphi) = r(\mathfrak{N},S,T) = O(h^k)$, $\mathrm{comp}(\Psi,S,T,\varepsilon) = O\left(\frac{c_1}{k\sqrt{\varepsilon}}\right)$.

Furthermore it is possible to show that for any distribution of points t_i in (13.2) there exists a positive constant a_k such that $r(\mathfrak{N},S,T) \ge a_k h^k$ for all $n \ge k$. Thus choosing the points equally spaced does not change the dependence on h.

We note that the solution operator S defined by (13.1) can describe a boundary value problem. Indeed, if G is the Green function of a boundary value problem then $\alpha(x) = (Sf)(x)$ is the solution of this problem. For instance if

(13.13) $$G(x,t) = \begin{cases} t(1-x) & 0 \le t \le x, \\ x(1-t) & x \le t \le 1, \end{cases}$$

then α satisfies the boundary value problem

(13.14) $$\begin{cases} \alpha''(x) = -f(x) & \text{for } x \in (0,1) \\ \alpha(0) = \alpha(1) = 0 . \end{cases}$$

14. EVALUATION OF DERIVATIVES

Let $\mathfrak{I}_1 = C^k(-1,1)$, $\mathfrak{I}_2 = C^{k-1}(-1,1)$ and $\mathfrak{I}_4 = C(-1,1)$ for an integer $k \geq 1$ with the uniform norm $\|f\| = \sup_{-1<t<1} |f(t)|$. Let

$$(14.1) \quad Sf = f'(0), \quad Tf = \frac{1}{k!} f^{(k)}$$

and let the information operator $\mathfrak{N}$ be given by

$$(14.2) \quad \mathfrak{N}(f;h) = [f(0), f(h), f(-h), \ldots, f(nh), f(-nh)]^t$$

where the parameter $h \in (0,1/n)$, $\text{card}(\mathfrak{N}) = 2n+1$.

Thus, for a given h we want to find an approximation to f'(0) knowing function evaluations at the points jh for $j = 0, \pm 1, \ldots, \pm n$. Note that round-off error analysis indicates that h should not be too small.

For the sake of simplicity we solve this problem for $k = 2n+1$. Note that $\ker \mathfrak{N} \cap T = \{0\} \subset \ker S$. Let

$$(14.3) \quad g_j(t) = \frac{\omega(t)}{(t-jh)\omega'(jh)} \text{ where } \omega(t) = t \prod_{i=1}^{n} (t^2 - i^2h^2)$$

for $j = 0, \pm 1, \ldots, \pm n$. Note that

$$(14.4) \quad g_0'(0) = 0, \quad g_j'(0) = \frac{(-1)^{j+1}(n!)^2}{jh(n+j)!(n-j)!}, \quad j = \pm 1, \ldots, \pm n.$$

Since $g_j(ih) = \delta_{ij}$ and $Tg_j = 0$, Lemma 8.3 assures us that

$$(14.5) \quad \varphi(\mathfrak{N}(f;h)) = \sum_{j=-n}^{n} f(jh) Sg_j(t) = \frac{(n!)^2}{h} \sum_{j=1}^{n} \frac{(-1)^{j+1}}{j(n+j)!(n-j)!} (f(jh) - f(-jh))$$

is a linear interpolatory optimal error algorithm. The algorithm (14.5) is

known as the nth central difference formula. We seek $e(\varphi)$. Since $f(t) - \sum_{j=-n}^{n} f(jh)g_j(t) = g(t)\omega(t)$ where $g(t)$ is the kth normalized divided difference of f and $|g(t)| \le 1$, then

$$f'(0) - \varphi(\mathfrak{N}(f;h)) = g(0)\omega'(0).$$

This yields

$$(14.6) \quad e(\varphi) = r(\mathfrak{N},S,T) = (n!)^2 h^{2n}.$$

Werschulz [77b] considers the dependence of $e(\varphi)$ on h and says φ has order of accuracy p if $e(\varphi) = O(h^p)$. The equation (14.6) agrees with his result that every algorithm which uses the information (14.2) has order of accuracy no greater than 2n.

We analyze the complexity of evaluation of derivatives. To find an ϵ-approximation to $f'(0)$ for every $f \in \mathfrak{I}_0$ we require

$$(14.7) \quad e(\varphi) = (n!)^2 h^{2n} < \epsilon.$$

With h, ϵ fixed this determines n. Note that h might be chosen as small as possible consistent with good round-off. If (14.7) holds then

$$(14.8) \quad (2n+1)(c_1+1) - 1 \le \mathrm{comp}(\mathfrak{N},S,T,\epsilon) \le (2n+1)(c_1+2) - 1$$

where c_1 is the complexity of one function evaluation.

15. NONLINEAR EQUATIONS

There exist problems for which we usually do not use fixed information $\mathcal{N}$. For instance, the solution of a nonlinear equation $f(x) = 0$ is often approximated by an iterative algorithm which repetitively uses an information operator $\mathcal{N}(f;x_i)$ where x_i, $i = 1,2,\ldots$, is a current approximation to the sought solution. Information operators of the form $\mathcal{N}(f,x_i)$ are studied in Traub and Woźniakowski [77c]. Here we want to show that it is also possible to deal with nonlinear equations for a fixed information operator. We shall show that some known iterations are "asymptotically" optimal error algorithms. Since this problem is nonlinear we cannot use the results of Sections 4 through 9. Our analysis will be based mostly on Section 2.

Let

$$(15.1)\quad f: D \subset B_1 \to B_2$$

where $D = \{x: \|x\| \leq 2R\}$ and B_1, B_2 are Banach spaces over the real or complex fields of dimension m, $m = \dim(B_1) = \dim(B_2)$, $1 \leq m \leq +\infty$. Let $\mathcal{F}_1$ be the class of all operators f which are k-times differentiable in the Frechet sense on D, $k \geq 2$. Define

$$(15.2)\quad \mathcal{F}_0 = \mathcal{F}_0(A_2,A_k) = \{f: f \in \mathcal{F}_1 \text{ and there exists } \alpha = \alpha(f), \|\alpha\| \leq R, \text{ such that } f(\alpha) = 0 \text{ and } \|f'(\alpha)^{-1}\frac{f''(x)}{2}\| \leq A_2 \text{ and } \|f'(\alpha)^{-1}\frac{f^{(k)}(x)}{k!}\| \leq A_k \text{ for all } x \in D\}$$

for constants A_2 and A_k which satisfy the condition

$$(15.3)\quad 2kA_k(3R)^{k-1} + 2A_2R < 1.$$

The solution of nonlinear equations in $\mathcal{F}_0$ is described by

$$(15.4)\quad S(f) = f^{-1}(0), \quad \mathcal{F}_2 = B_2.$$

We first show that S is well-defined. It suffices to prove that $f(x) = 0$ for $\|x\| \le R$ has a unique solution $\alpha = f^{-1}(0)$, for $f \in \mathcal{F}_0$. Let

$$(15.5)\quad R_j(x,y;f) = \int_0^1 f^{(j)}(y + t(x-y))(x-y)^j \frac{(1-t)^{j-1}}{(j-1)!}\, dt$$

for $x,y \in D$ and $j \le k$. Then

$$(15.6)\quad f(x) = f(\alpha) + f'(\alpha)(x-\alpha) + R_2(x,\alpha;f)$$

and $f(x) = 0$ is equivalent to the equation

$$x-\alpha = -f'(\alpha)^{-1} R_2(x,\alpha;f).$$

From (15.2) and (15.3) we get for $\|x\| \le R$,

$$\|x-\alpha\| \le A_2 \|x-\alpha\|^2 \le A_2 R \|x-\alpha\| \le \frac{1}{2}\|x-\alpha\|$$

which implies $x = \alpha$. Thus S is well-defined by (15.4) and $\alpha = S(f)$ satisfies the nonlinear equation $f(\alpha) = 0$.

Define the information operator

$$(15.7)\quad \mathcal{N}(f) = [y(f), f(y(f)), \ldots, f^{(k-1)}(y(f))]^t$$

where $y = y(f)$ is an approximation to the solution $\alpha = S(f)$, $\|y\| \le R$.

We want to find $d(\mathcal{N},S)$, the diameter of information $\mathcal{N}$ for the problem S. See (2.9).

Lemma 15.1

$$(15.8) \qquad d(\mathfrak{N},S) \le \frac{2A_k}{1-A_2R} \sup_{f\in\mathfrak{I}_0} \left(\frac{3}{2}\| y(f)-S(f)\|^k\right).$$

Proof

Note that $\mathfrak{N}(\tilde{f}) = \mathfrak{N}(f)$ implies $\tilde{f}^{(j)}(y) = f^{(j)}(y)$ for $j = 0,1,\ldots,k-1$ and $\tilde{f}$, $f \in \mathfrak{I}_0$. Then

$$f(x)-\tilde{f}(x) = R_k(x,y,f-\tilde{f}).$$

Since $\tilde{f}(x) = 0$ is equivalent to $f(x) = R_k(x,y,f-\tilde{f})$ and f satisfies (15.6), we get

$$(15.9) \quad x = H(x) \overset{df}{=} \alpha + f'(\alpha)^{-1}\{R_k(x,y,f-\tilde{f}) - R_2(x,\alpha,f)\}.$$

We show that H is a contraction on $J = \{x: \|x-\alpha\| \le \frac{1}{2}\|y-\alpha\|\}$. Indeed, $\|H(x)-\alpha\| \le 2A_k\|x-y\|^k + A_2\|x-\alpha\|^2 \le 2A_k(\frac{3}{2})^k\|y-\alpha\|^k + \frac{1}{4}A_2\|y-\alpha\|^2 \le \frac{1}{2}\|y-\alpha\|(6A_k(3R)^{k-1} + A_2R) \le \frac{1}{2}\|y-\alpha\|$ due to (15.2) and (15.3) Furthermore

$$\|H'(x)\| \le 2kA_k\|x-y\|^{k-1} + 2A_2\|x-\alpha\| \le 2kA_k(2R)^{k-1} + 2A_2R < 1$$

due to (15.3). Thus the equation (15.9) has a unique solution $\tilde{\alpha}$, $\|\tilde{\alpha}\| \le \frac{1}{2}\|y-\alpha\| \le R$. Set $x = \tilde{\alpha}$ in (15.9). Then $\|\tilde{\alpha}-\alpha\| \le 2A_k\|\tilde{\alpha}-y\|^k + A_2\|\tilde{\alpha}-\alpha\|^2$ which yields

$$\|\tilde{\alpha}-\alpha\| \le \frac{2A_k(\frac{3}{2}\|y-\alpha\|)^k}{1-A_2R}.$$

This proves (15.8) and completes the proof. ■

We want to prove that (15.8) is, in general, sharp with respect to $\|y(f)-S(f)\|^k$.

Lemma 15.2

If y(f) approaches $\alpha = S(f)$ then

(15.10) $\operatorname{diam}(\mathfrak{N},S) = 2A_k \sup_{f\in\mathfrak{I}_0} \{\| y(f)-S(f)\|^k (1 + o(\| y(f)-S(f)\|))\}.$ ■

Proof

The equation (15.9) for $x = \tilde{\alpha}$ yields

(15.11) $\| \tilde{\alpha}-\alpha\| \le 2A_k \| y(f)-\alpha\|^k + A_2 \| \tilde{\alpha}-\alpha\|^2.$

Since $\| \tilde{\alpha}-\alpha\| = O(\| y(f)-\alpha\|^k)$, (15.11) can be rewritten as $\| \tilde{\alpha}-\alpha\| \le 2A_k \| y(f)-\alpha\|^k (1 + o(1))$. Since this bound is sharp we've proven (15.10). ■

Lemma 10.2 states that the diameter $\operatorname{diam}(\mathfrak{N},S)$ is roughly equal to $2A_k \sup_{f\in\mathfrak{I}_0} \| y(f)-S(f)\|^k$ where k is the first omitted derivative in the information (15.7).

We establish asymptotically optimal error algorithms for the problem S. Let

(15.12) $\tilde{f}(x) = f(y) + f'(y)(x-y) + \ldots + \frac{1}{(k-1)!} f^{(k-1)}(y)(x-y)^{k-1}.$

Note that $\mathfrak{N}(\tilde{f}) = \mathfrak{N}(f)$. From (3.10) in Traub and Woźniakowski [77b] we know that

$$\| \tilde{f}'^{-1}(\tilde{\alpha})^{-1} \frac{\tilde{f}''(x)}{k!}\| \le \frac{A_2 + \frac{k(k-1)}{2}A_k(2\| y-\alpha\|)^{k-2}}{1 - A_2\| y-\alpha\| - kA_k(\frac{3}{2}\| y-\alpha\|)^{k-1}} \stackrel{df}{=} \tilde{A}_2(y).$$

Thus $\tilde{f} \in \mathfrak{I}_0(\tilde{A}_2(y),0)$ where $\tilde{A}_2(y) = A_2 + O(\| y-\alpha\|)$. Define the algorithm

(15.13) $\varphi(\mathfrak{N}(f)) = S(\tilde{f}),$

i.e., $\varphi(\mathfrak{N}(f))$ is a unique solution of the nonlinear equation $\tilde{f}(x) = 0$. The algorithm (15.13) is known as the interpolatory iteration I_n and was considered by Traub and Woźniakowski [76b, 77a, 77b]. Note that for n = 2 we

get one step of Newton iteration since $\tilde{f}(x) = f(y) + f'(y)(x-y)$ and $\varphi(\mathfrak{N}(f)) = \tilde{\alpha} = f(y) - f'(y)^{-1}f(y)$.

Lemma 15.3

(i) For any $y(f)$ the error of φ is bounded by

$$(15.14)\quad e(\varphi) \le \frac{A_k}{1-A_2R} \sup_{f\in\mathfrak{F}_0} \left(\frac{3}{2}\| y(f)-S(f) \|\right)^k.$$

(ii) If $y(f)$ approaches $\alpha = S(f)$ then the algorithm φ is asymptotically optimal, i.e.,

$$(15.15)\quad e(\varphi) \cong r(\mathfrak{N},S) \cong A_k \sup_{f\in\mathfrak{F}_0} \| y(f)-S(f) \|^k. \qquad \blacksquare$$

Proof

To prove (15.14) we repeat the proof of Lemma 10.1. Note that $R_k(x,y,f-\tilde{f})$ in (15.9) for $\tilde{f}$ defined by (15.12) has the bound $\| R_k(x,y,f-f) \| \le A_k \| x-y \|^k$ which yields (15.14). The same argument enables us to conclude that $e(\varphi) \cong A_k \sup_{f\ \mathfrak{F}_0} \| y(f)-S(f) \|^k$ for $y(f)$ approaching α. Thus, from Lemma 15.2 we get

$$e(\varphi) \cong \frac{1}{2}\, d(\mathfrak{N},S) \cong r(\mathfrak{N},S)$$

which proves (15.15). $\blacksquare$

The algorithm (15.13) is known to have maximal order of convergence among all iterations using the information of (15.7); see Traub and Woźniakowski [76_a]. Lemma 15.3 states that this algorithm has asymptotically optimal error in the class $\mathfrak{F}_0$.

Complexity of the algorithm (15.13) and its dependence on k were considered in detail by Traub and Woźniakowski [77b].

CHAPTER IV

CONCLUDING REMARKS

16. COMMENTS AND EXTENSIONS

This is the first of a series of papers in which we develop an information based theory of optimal error algorithms and of problem complexity. We conclude this paper by a partial list of interesting problems and extensions which will be studied in the future.

1. In this paper we restrict ourselves to general information operators $\mathfrak{N} = \mathfrak{N}(f)$ and illustrate our concepts and results by simple examples. Future papers will be devoted to complete analysis of an application area in this general framework. We will include the effect of computing $\mathfrak{N}(f)$ approximately, problem condition, algorithm stability, and the cost of arithmetic precision.

2. For some problems information operators have additional structure. For instance, "iterative" information operators $\mathfrak{N} = \mathfrak{N}(f,x_i,x_{i-1},\ldots,x_{i-m})$ depend on current approximations to the solution and they are used for $i = m,m+1,\ldots$. The parameter m measures the size of "memory" used in the process. For $m = 0$ we have information operators "without memory", for $m \geq 1$ "with memory". See Traub and Woźniakowski [77d].

3. In Chapter 2 we deal with linear information mostly of the form $\mathfrak{N}(f) = [L_1(f),L_2(f),\ldots,L_n(f)]^t$ defined by n independently given linear functionals $L_1,L_2,\ldots,L_n$. A natural generalization is an "adaptive" linear information operator of the form

$$\mathfrak{N}(f) = [L_1(f), L_2(f;L_1(f)), \ldots, L_n(f;L_1(f), \ldots, L_{n-1}(f))]^t$$

where L_i depends linearly on its first argument. This form enables us to use the previously computed functionals to determine the next functional. Such adaptive information is widely used in practice in a number of application areas.

4. In Chapter 2 we deal with linear problems defined by two linear operators S and T. The restriction operator T defines the domain of problem elements. In several applications we take $T = \frac{1}{k!} D^k$, the kth normalized derivative operator, and we show that complexity decreases as the regularity parameter k increases. We believe this to be a general phenomenon.

5. For some problems the domain $\mathfrak{I}_0$ is defined by a two sided inequality on T, i.e., $\mathfrak{I}_0 = \{f \in \mathfrak{I}_1 : 0 < c \leq \|Tf\| \leq 1\}$ for a constant c. Sometimes $\mathfrak{I}_0$ is defined by more than one linear operator, e.g., $\mathfrak{I}_0 = \{f \in \mathfrak{I}_1 : c_i \leq \|T_i f\| \leq 1, \text{ for } i = 1,2,\ldots,m\}$ for nonnegative constants $c_1, c_2, \ldots, c_m$.

6. All basic results of Chapter 2 are for a linear operator S. Much of the linear theory can be generalized to nonlinear operators S. For instance, assume that there exist two linear operators S_1 and S_2 such that

$$\|S_1(f_1-f_2)\| \leq \|S(f_1)-S(f_2)\| \leq \|S_2(f_1-f_2)\|$$

for all f_1 and f_2 from the domain of S. Then

$$d(\mathfrak{N},S_1) \leq d(\mathfrak{N},S) \leq d(\mathfrak{N},S_2)$$

and one can apply the linear theory to S_1 and S_2 to derive lower and upper bounds on the diameter $d(\mathfrak{N},S)$ and complexity of the problem S.

ACKNOWLEDGMENTS

We wish to thank J. Bentley, A. Bojańczyk, H. T. Kung, M. Sapsford, M. Shamos, B. Weide, and A. Werschulz for their comments on the manuscript. We are grateful to Dorothy Josephson for her accurate and speedy typing.

GLOSSARY

We summarize below basic concepts used throughout the paper. We list a symbol, its meaning, and the section reference where this symbol appears for the first time.

Symbol	Meaning	Section Reference
S	the solution operator, sometimes called the problem, $S: \mathfrak{I}_0 \to \mathfrak{I}_2$ and $\mathfrak{I}_0 \subset \mathfrak{I}_1$	2, (2.1)
$\mathfrak{I}_0$	the domain of S	2, (2.1)
$\mathfrak{I}_1$	linear space, $\mathfrak{I}_0 \subset \mathfrak{I}_1$	2
$\mathfrak{I}_2$	the range of S	2
ϵ	error parameter, $\epsilon > 0$	2
$x = x(f)$	ϵ-approximation, $\|x-\alpha\| < \epsilon$,	2, (2.2)
f	the problem element, $f \in \mathfrak{I}_0$	2
α	the solution element $\alpha = S(f)$	2
$\mathfrak{N}$	the information operator, $\mathfrak{N}: D_{\mathfrak{N}} \to \mathfrak{I}_3$	2, (2.3)
$\mathfrak{I}_3$	the range of $\mathfrak{N}$	2, (2.3)
$d(\mathfrak{N},S)$	the diameter of information $\mathfrak{N}$ for the problem S	2, (2.9)
$r(\mathfrak{N},S)$	the radius of information $\mathfrak{N}$ for the problem S	2, (2.10)
φ	algorithm, $\varphi: \mathfrak{N}(\mathfrak{I}_0) \to \mathfrak{I}_2$	2
$e(\varphi)$	the error of algorithm φ	2, (2.13)
$\Phi(\mathfrak{N},S)$	the class of all algorithms using the information $\mathfrak{N}$ for the problem S	2
φ^I	interpolatory algorithm	2, (2.16)
$e(\mathfrak{N},S)$	the optimal error	2, (2.18)
φ^{oe}	optimal error algorithm	2, (2.19)
φ^c	central algorithm	2, (2.23)

Symbol	Meaning	Section Reference
P	the set of primitives	3
$\text{comp}(\mathfrak{N}(f))$	the information complexity of computing $\mathfrak{N}(f)$ where $\mathfrak{N}$ is a permissible information operator	3
$\text{comp}(\varphi(y))$	the combinatory complexity of computing y where φ is a permissible algorithm	3
$\Phi(\epsilon)$	the class of all permissible algorithms for which $e(\varphi) < \epsilon$	3
$r(\mathfrak{N},S) \geq \epsilon$	the problem S with information $\mathfrak{N}$ is ϵ-non-computable	3
$r(\mathfrak{N},S) < \epsilon$	the problem S with permissible $\mathfrak{N}$ and $\Phi(\epsilon) \neq \emptyset$ is ϵ-computable with respect to P	3
$\text{comp}(\varphi)$	the complexity of an algorithm φ	3, (3.1)
$\text{comp}(\mathfrak{N},S,\epsilon)$	the ϵ-complexity of the information $\mathfrak{N}$ for the problem S	3, (3.2)
φ^{oc}	optimal complexity algorithm	3, (3.3)
$\text{comp}(\mathfrak{N})$	the information complexity	3, (3.4)
Ψ	a class of permissible information operators	3, (3.8)
$\text{comp}(\Psi,S,\epsilon)$	the ϵ-complexity of the problem S in the class Ψ	3, (3.9)
$\mathfrak{N}_1 \subset \mathfrak{N}_2$	$\ker \mathfrak{N}_2 \subset \ker \mathfrak{N}_1$	4, def. 4.1
$\mathfrak{N}_1 \asymp \mathfrak{N}_2$	$\ker \mathfrak{N}_1 = \ker \mathfrak{N}_2$	4, def. 4.1
$A^{\perp}$	algebraic complement of A	4, (4.2)
$\text{codim } A$	codimension of A	4, (4.2)
$\text{card}(\mathfrak{N})$	the cardinality of the information $\mathfrak{N}$	4, (4.5)
T	the restriction operator, $T: \mathfrak{I}_1 \to \mathfrak{I}_4$	5, (5.1)
$\mathfrak{I}_4$	the range of T	5, (5.1)
$d(\mathfrak{N},S,T)$	the diameter of information $\mathfrak{N}$ for the problem (S,T)	5
$\text{index}(S,T)$	the index of the problem (S,T)	5, def. 5.1

Symbol	Meaning	Section Reference
$A(S,T)$	algebraic complement of ker T $\cap$ ker S in the space ker T	5, (5.3)
$\xi_1^*,\ldots,\xi_{n^*}^*$	basis of $A(T,S)$, $n^* = \text{index}(S,T)$	5, (5.3)
$\mathfrak{N}^*$	information operator such that $\text{card}(\mathfrak{N}^*) = \text{index}(S,T)$ and $\ker \mathfrak{N}^* \cap \ker T \subset \ker S$	6, (6.1)
Ψ_n	the class of all information operators $\mathfrak{N}$ such that $\mathfrak{N}^* \subset \mathfrak{N}$ and $\text{card}(\mathfrak{N}) \leq n$	6
T^{-1}	the inverse operator of T	6, (6.3)
$d(n,S,T)$	the n-th minimal diameter of information	6, (6.6)
$\mathfrak{N}_n^{oi}$	n-th optimal information	6, (6.7)
K	the linear operator $K = ST^{-1}$	6, (6.9)
$b(m,K)$	the m-th minimal norm of the linear operator K	6, (6.9)
B_m	m-th minimal subspace of the linear operator K	6, (6.11)
$d(S,T)$	the diameter of problem error	6, def. 6.2
$d(S,T) = +\infty$	the problem (S,T) is strongly non-computable	6, def. 6.2
$d(S,T) \geq 2\epsilon$	the problem (S,T) is ϵ-non-computable	6, def. 6.2
$d(S,T) = 0$	the problem is convergent	6, def. 6.2
$m(\Psi,S,T,\epsilon)$	the ϵ-cardinality number for the problem (S,T) in the class Ψ	9, (9.5)
Ψ_U	the class of all linear information operators $\mathfrak{N}$ such that $\text{card}(\mathfrak{N}) < +\infty$	9

BIBLIOGRAPHY

Akhieser [56] — Akhieser, N. I., Theory of Approximation, Frederick Ungar Publishing Co., New York, 1956.

Bakhvalov [71] — Bakhvalov, N. S.,"On the Optimality of Linear Methods for Operator Approximation in Convex Classes of Functions,"USSR Computational Mathematics and Mathematical Physics 11, 1971, 244-249.

Bojanov [74] — Bojanov, B. D., "Best Quadrature Formula for a Certain Class of Analytic Functions," Zastosowania Matematyki, XIV, 1974, 441-447.

Bojanov [76] — Bojanov, B. D., "Optimal Methods of Integration in the Class of Differentiable Functions," Zastosowania Matematyki, XV, 1976, 105-115.

Borodin [72] — Borodin, A., "Complexity Classes of Recursive Functions and the Existence of Complexity Gaps," J.ACM 19, 1972, 158-174. Corrigendum, J.ACM 19, 1972, p. 576.

Brent, Winograd and Wolfe [74] — Brent, R. P., Winograd, S. and Wolfe, P., "Optimal Iterative Processes for Root-finding," Num. Math. 20, 1973, 327-341.

Garey and Johnson [76] — Garey, M. R. and Johnson, D. S., "Approximation Algorithms for Combinatorial Problems: An Annotated Bibliography," in Algorithms and Complexity: New Directions and Recent Results, edited by J. F. Traub, Academic Press, 1976, 41-52.

Golomb and Weinberger [59] — Golomb, M. and Weinberger, H. F., "Optimal Approximation and Error Bounds," in On Numerical Approximation, edited by R. E. Langer, The University of Wisconsin Press, 1959, 117-190.

Kacewicz [75] — Kacewicz, B., "Integrals with a Kernel in the Solution of Nonlinear Equations in N Dimensions," Department of Computer Science Report, Carnegie-Mellon University, 1975.

Kacewicz [76a] — Kacewicz, B., "An Integral-Interpolation Iterative Method for the Solution of Scalar Equations," Numer. Math. 26, 1976, 355-365.

Kacewicz [76b] — Kacewicz, B., "The Use of Integrals in the Solution of Nonlinear Equations in N Dimensions," in Analytic Computational Complexity, edited by J. F. Traub, Academic Press, 1976, 127-141.

Kung [76] — Kung, H. T., "The Complexity of Obtaining Starting Points for Solving Operator Equations by Newton's Method," in Analytic Computational Complexity, Academic Press, 1976, 35-59.

Meersman [76a] — Meersman, R., "Optimal Use of Information in Certain Iterative Processes," in Analytic Computational Complexity, Academic Press, 1976, 109-127.

Meersman [76b] — Meersman, R., "On Maximal Order of Families of Iterations for Nonlinear Equations," Doctoral Thesis, Vrije Universiteit Brussel, Brussels, 1976.

Micchelli and Rivlin [77] — Micchelli, C. A. and Rivlin, T. J., "A Survey of Optimal Recovery," in Proceedings of the International Conference of Optimal Estimation in Approximation Theory, to appear.

Micchelli, Rivlin and Winograd [76] — Micchelli, C. A., Rivlin, T. J. and Winograd, S., "Optimal Recovery of Smooth Functions," Numer. Math. 26, 1976, 191-200.

Schultz [73] — Schultz, M. H., "The Complexity of Linear Approximation Algorithms," in Complexity of Computation, edited by R. M. Karp, AMS 1974, 135-149.

Smolyak [65] — Smolyak, S. A., "On an Optimal Restoration of Functions and Functionals of Them," Candidate Dissertation, Moscow State University, 1965.

Tihomirov [69] — Tihomirov, V. M., "Best Methods of Approximation and Interpolation of Differentiable Functions in the Space C[-1,1]," Math. USSR Sbornik, 9, 1969, 275-289.

Traub and Woźniakowski [76a] — Traub, J. F. and Woźniakowski, H., "Optimal Linear Information for the Solution of Non-Linear Equations," in Algorithms and Complexity: New Directions and Recent Results, Academic Press, 1976, 103-119.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* GENERAL THEORY OF OPTIMAL ERROR ALGORITHMS AND ANALYTIC COMPLEXITY		5. TYPE OF REPORT & PERIOD COVERED Interim 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) J. F. Traub and H. Wozniakowski		8. CONTRACT OR GRANT NUMBER(s) ,MCS75-222-55 N00014-76-C-0370, NR 044-422
9. PERFORMING ORGANIZATION NAME AND ADDRESS Carnegie-Mellon University Computer Science Dept. Pittsburgh, PA 15213		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE September 1977 13. NUMBER OF PAGES 95
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)* This is the first of a series of papers constructing an information based general theory of optimal errors and analytic computational complexity. Among the applications are such traditionally diverse areas as approximation, boundary-value problems, quadrature, and nonlinear equations in a finite or infinite dimensional space. Traditionally algorithms are often derived by ad hoc criteria. The information based theory rationalizes the synthesis of algorithms by showing how to construct algorithms which minimize or nearly minimize the error. For certain classes of problems it shows how to construct algorithms (linear optimal error algorithms)

continued

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

20. abstract (continued)

which enjoy essentially optimal complexity with respect to all Possible algorithms. The existence of "strongly non-computable" problems is demonstrated. In contrast with the gap theorem of recursively computable functions we show that "every monotonic" real function is the complexity of some problem.